A Monte Carlo Simulation Study of the Factors Influencing the Performance of Flood Early Warning Systems

Luis Felipe Duque Yaguache

A thesis presented for the degree of Doctor of Philosophy.

Supervisors:

Prof. Enda O'Connell

Dr. Greg O' Donnell

School of Engineering

December 2020

Declaration of Originality

I hereby certify that the work presented in this thesis is my original research work **with the exception of the calibration of the RAINSIM model and generation of storms and their forecasts in Chapter 7, sections 7.3.1 -7.3.3, which was carried out by Dr. G. O'Donnell.** Due reference is given to literature and any research collaborations where appropriate. The research presented in this thesis has not been submitted for any other degree or professional qualification.

Luis Felipe Duque Yaguache

Abstract

In recent decades, flood early warning systems (FEWSs) have been widely used as complementary non-structural mitigation measures in order to improve the population resilience to floods. FEWS research focusses mainly on flood forecasting techniques or social aspects of warning response, and end-to-end modelling frameworks that represent the entire FEWS forecast-decision-response/impact chain are rarely developed.

A generic Monte Carlo simulation framework has been developed that represents an end-to-end FEWS in a versatile way, allowing factors influencing FEWS performance to be explored which cannot be analysed easily based on limited real-world data. The framework has been applied to a simulated generic fluvial case, where factors influencing FEWS performance in terms of reliability and economic effectiveness are explored. A new reliability performance measure based on inundation maps has been proposed. The framework has also been used to explore factors controlling the performance of a simulated FEWS representing an urban polder in Nanjing, China, with performance metrics based on waterlogging and pumping costs.

For the generic fluvial case, the main results show that: i) the correlation between forecasts and observed values controls reliability; ii) probabilistic forecasts based on optimising a probabilistic threshold are robust to forecast biases in the mean and variance, iii) a FEWS based on uncertain forecasts is characterised by an optimal lead time that represents a balance between an adequate time to act in response and a reasonably good forecast; iv) the performance of the proactive action is the most important factor influencing the economic effectiveness of a FEWS. For the simulated flood-prone polder system case study, the results show that probabilistic forecasts of storm rainfall and runoff volume can considerably enhance the waterlogging and pumping metrics.

The results of this research can be used to improve the performance of fluvial FEWSs, and to design FEWSs for polder systems.

Dedication

I dedicate this thesis lovingly to a new angel watching over me, my father,

Edwin Bolivar+

Acknowledgements

I am grateful to a large number of people who have contributed both personally and professionally to the completion of this research. First and foremost, I would like to thank my supervisors, Prof. Enda O'Connell and Dr. Greg O' Donnell, for their academic support, mentoring, and guidance throughout this project. Thanks for taking me on as a student and for giving me the opportunity to develop this research. Your invaluable insights and advice have all been very much appreciated.

Financial support for this research was provided through a Ph.D. scholarship from the Secretary for Higher Education, Science, Technology and Innovation (SENESCYT) of the Government of Ecuador. Thanks also to my family in Ecuador, who have been with me every step of the way – thanks, Mom and Brother.

Finally, I could not have completed this degree without the unconditional love and continuous support of my wife, Viviana, who has been so patient and encouraging during the course of this research.

Luis Felipe Duque Yaguache

Contents

List of Figures

List of Tables

Chapter 1. Introduction

1.1 Background

Climate change and the intense urbanization of flood-prone areas evidenced in the last decades have altered the hydrological responses of catchments, increasing the frequency and magnitude of floods in many countries across the globe. As existing standards of protection based on structural measures are proving to be inadequate under climate change, and the concept of flood protection has given way to that of flood risk management, effective and reliable flood early warning systems (FEWSs) are becoming vital complementary non-structural mitigation measures in order to improve the population resilience to this natural hazard (WMO, 2011; Milionis and Owen, 2018). As a result, a new science, known as Flood Warning Science, which covers a wide range of disciplines ranging from hydrometeorological science to social psychology, has emerged from the need to assess, research, and improve the process chain characterizing these systems (Parker, 2017).

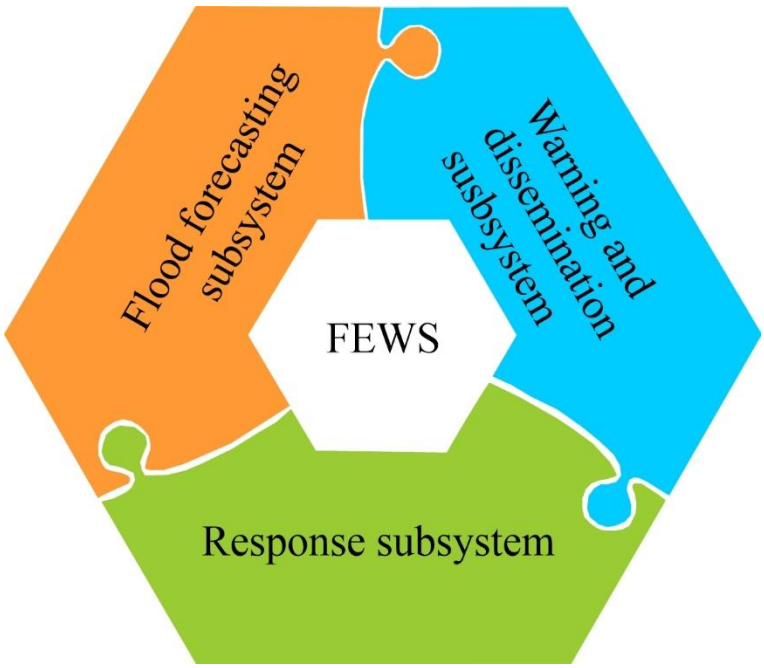


Figure 1.1. Illustration of a Flood Early Warning System (FEWS)

A FEWS aims to provide timely and meaningful warning information to people threatened by floods to allow them to act in advance to reduce the risks involved. This system can be represented by three main subsystems (Parker and Fordham, 1996; Carsell *et al.*, 2004): i) a flood forecasting subsystem, ii) a warning and dissemination subsystem, and iii) a response subsystem (Figure 1.1).

The flood forecasting subsystem is the most well-recognized and studied element and uses a forecasting model to generate forecasts on which warnings are based. These forecasts can be generated by, for example, conceptual or physically-based or rainfall-runoff and/or flooding routing models, such as the Stanford Watershed Model (Crawford and Linsley, 1966), ARNO (Todini, 1996), or TOPKAPI (Liu and Todini, 2002) models, among others. Advanced data-driven models, e.g., machine learning models based on fuzzy systems, artificial neural networks, or support vector regression, are also used for these purposes (Mosavi *et al.*, 2018).

The forecasting model's inherent uncertainty means that flood warnings are based on uncertain information (Basher *et al.*, 2006). The quantification and reduction of this uncertainty (*hereafter called predictive uncertainty*) are the main challenges faced by forecasters (Perera,2020). Despite the significant research progress made in the last two decades on this topic, operational flood warnings have been largely based on deterministic forecasts, though probabilistic forecasting is starting to be considered in practice (see the English case in Arnal *et al.*, (2020)).

One of the reasons for the slow uptake of probabilistic forecasting can be the confusion in the literature over the appropriate measure of uncertainty to use in flood warning. Predictive Uncertainty (PU) is the appropriate measure to use (Krzysztofowicz, 1999a; Todini, 2016). Todini (2008) defines PU as the predictive density function of the unknown future quantity, typically the observed peak water level or discharge at a specific cross-section, conditional on the forecast of that level for a given lead time (τ). To introduce the reader to this concept without the formal mathematical definitions that come later, the example shown in Figure 1.2 is used. Deterministic and probabilistic forecasts of the water levels for a lead time of $\tau = 6$ hours are shown Figure 1.2a and 1.2b.

To understand the concept of PU, the following question is posed: flood damage in Figure 1.2a will occur: (1) when the *forecasted water level* overtops the dykes, or (2) when the *actual future water level* overtops the dykes? The obvious answer is (2), i.e., when the *actual future water level* overtops the dykes. Therefore, the model forecast must not be considered as a *reality* but just as a *virtual reality*, which, however, contains essential information to reduce our uncertainty on what will actually occur. PU is, thus, the uncertainty about a *real future value* (**not** the model forecast) i.e., the water level to be observed in 6 hours from now, conditional upon

the model forecast (Figure 1.2b). Knowing the probability distribution of the future water level conditional on the forecast value which is the formal definition of PU, one can estimate the probability of exceedance (PE) of the flooding level (red area in Figure 1.2b), and make a warning decision based on this (Todini, 2010). If the value of PE exceeds a predefined probability threshold PT, a warning is issued. PU, PE and PT are used throughout this thesis to inform simulated flood warning decisions, and any proactive actions taken in response to reduce flood damage. The results are compared with those from deterministic forecasting.

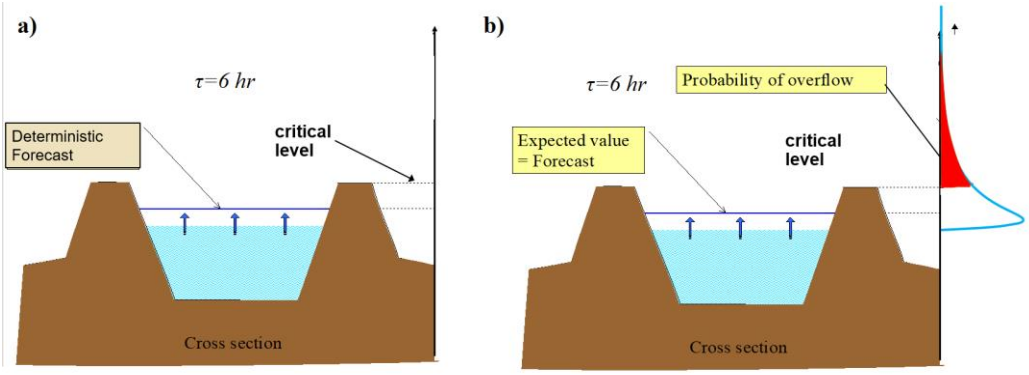


Figure 1.2. Illustration of the PU concept
Adapted from Todini (2004) and Todini (2010)

On the other hand, the warning and dissemination subsystem and the response subsystem have not received the same attention as the flood forecasting subsystem (Figure 1.1). The former involves all the processes needed to make the flood warnings reach those at risk, whereas the latter component has to do with the capacity of an at-risk community to take actions that mitigate the flood impact (Molinari *et al.*, 2013). These elements cover vulnerability factors, which are important to reduce flood risk, and are as important as the flood forecasting subsystem as it has been proved that a high-quality forecast is not enough to reduce flood impact (Basher *et al.*, 2006).

Although it is not always the case, there is usually a real-time information flow among components of the FEWS. This information can be either intrinsic, when it is part of the system and incorporate formal warnings, or extrinsic, which involves informal warnings used by floodplain users (Parker and Fordham, 1996). Therefore, the success of the FEWS heavily depends on how effective the communication is among the personnel involved in each component.

1.2 Research needs

Many studies have focused on improving individual components of a FEWS, such as forecasts or social vulnerability. These studies often imply that an improvement to a component of the system suggests an improvement in the performance of the integrated system. Thus, in the literature, one can find many studies that include only the technical components (often related to flood forecasting techniques) or social components (often related to the warning response) of a FEWS. End-to-end modelling frameworks that analyse the performance of a full FEWS chain of components (Figure 1.1) are, thus, rarely developed. Such frameworks should allow forecast outputs to be linked with warning decisions and the associated responses and flood impacts and, at the same time, evaluate the main factors influencing the performance of the integrated system encompassing the forecast, warning, and response processes. A FEWS is only as good as its weakest link, which such an evaluation should help identify.

The end-to-end simulation of a FEWS is no easy task. Even though FEWSs started to be studied between the 1960s and 1970s (Williams, 1964; McLuckie, 1970; Mileti and Beck, 1975), it was not until the 1980s that Krzysztofowicz and Davis (1983) proposed a simulation framework to represent this system. In this work, the FEWS chain is reduced to a flood forecast-response process, where a forecast system is viewed as an information system providing forecasts to a response system, which is, in turn, viewed as a decision system where floodplain dwellers make their own decisions about evacuation based on a sequence of forecasts they receive. Krzysztofowicz (1993) also proposed a methodology to represent the FEWS until the warning component based on a monitor-forecast-decision chain. This framework is used to evaluate the performance of a FEWS, where Bayesian decision theory is used to represent warning decisions by a rule that maximizes the expected utility. The flood warning reliability and economic benefits are used to define the FEWS performance.

Verkade and Werner (2011) proposed a methodology to compare the benefits of a deterministic FEWS with its probabilistic version based on the relative economic value theory. In their methodology, the forecast-warning-decision chain is used to obtain the relative economic value (REV) in terms of hits, false alarms, and misses. REV is subsequently used to scale between the flood risk of two benchmark

scenarios: the no warning and perfect forecast scenarios. The flood risks of these benchmark scenarios are computed through the response and impact components of the framework. In this work, real-world data and modelling are used to represent the flood forecasting and the response/impact component. This work was the first attempt to include the economic consequences of predictive uncertainty, PU, in the economic effectiveness of a FEWS. However, this methodology assumes that the cost of the warning response is independent of the magnitude of the forecast flooding, and it only considers the economic consequences of "wrong" decisions (false alarms or misses) and negates the fact that the PU can also (negatively) impact the economic effectiveness of flood warnings after a "good decision", i..e, a flood preceded by a warning (hit), due to the flood magnitude's uncertainty. Thus, there is a need to explore the negative impact of the economic consequences of PU in a FEWS in all possible situations, i.e., hits, misses, and false alarms situations. Other restrictive assumptions limit the applicability of this framework.

Girons Lopez *et al.* (2017) also considered the end-to-end modelling of a FEWS. They used a Monte Carlo analytical framework to represent a FEWS through a forecast-decision-response-and-impact chain. Based on this framework, they explored the impact of social preparedness on the economic effectiveness of a FEWS and how it is related to the flood warning reliability. A strong limitation of this study is that it does not consider the impact of the lead time on the FEWS performance, which is an important factor to be considered in this type of analysis.

Bischiniotis *et al.* (2019) also used a forecast-decision-response-and-impact chain to represent a FEWS. Based on this chain, they developed a framework to analyse the economic value of a FEWS in which actions in the response component can be taken at different time points. The economic value is obtained based on the economic consequences of hits, misses, and false alarms that characterise the flood warning reliability of the system. Like Verkade and Werner (2011), this work also considered limited real-world data to represent the flood forecasting component; therefore, they are not able to base their results by considering a considerable range of potential flood events *and* warning decision situations that the area monitored by the FEWS may be subject to.

None of the works mentioned above uses the forecast-decision-response-and-impact chain to evaluate the flood warning reliability in terms of affected houses,

i.e., warned and flooded houses. With this information, one can evaluate the reliability of flood warnings based on whether a warned property was or not subsequently flooded. It is a more reliable approach than those used by these research works and other related works, which is based on whether a warning preceded a flood event's occurrence or non-occurrence in the floodplain *or* in a specific flood risk zone, where a pre-defined threshold defines the occurrence of the event. This common approach does not consider the flood magnitude's uncertainty. Note that, due to PU, there is always a difference between the magnitude of the forecast and observed flooding and, therefore, a difference between the warned and flood properties. Moreover, this approach does not exploit the information provided by inundation maps which are now become a routine element of the forecast component of a FEWS.

The author of the thesis could also identify that there is no generic framework to underpin the end-to-end simulation of a FEWS. Note that beyond each specific purpose, the research works mentioned above have idealized the FEWS through the forecast, warning, and response/impact/components. However, they have used ad-hoc frameworks that have been designed for a specific flood type (often river floods) or a specific response action using restrictive analytical assumptions. In this sense, there is a need to design a generic framework that defines well the main generic components involved in the end-to-end simulation of a FEWS and the type of information generated and used by each of them during the simulation of a flood threat. Also, to avoid the limitations imposed by real-world data, such as a limited time span or a small number of registered extreme events, this generic framework should consider synthetic observed and forecast information that is based on the properties of real data. This would parallel the seminal work performed by Hosking and Wallis (1988, 1997), who demonstrated how Monte Carlo sampling can be used to explore the sensitivity of the estimate of T-year flood to various factors and assumptions, leading to more robust new approaches. This generic framework should underpin the end-to-end simulation of a FEWS monitoring and warning any type of flood-prone area.

The author of the thesis could not identify any research works that use an end-to-end modelling framework to explore the performance of a FEWS monitoring and warning a flood-prone polder system. No research works could be identified reporting on either the design, implementation, or evaluation of FEWSs operating

for flood-prone polder areas. Polders are areas that lie below the levels of the sea or adjacent outer rivers, and, therefore, pumping systems must be used during storm events to remove water from the inner rivers (artificial rivers inside the polder area) to the outer rivers to enable water to drain from the polder areas into the inner rivers. In a flood-prone polder area, the pumping capacity is invariably lower than the drainage capacity (maximum runoff entering the inner rivers). Since the polder system's storage capacity is defined by the water level of the inner rivers when the storm arrives, flood warnings can provide time in advance to decrease that water level (increase the storage capacity) and avoid the storage capacity being overwhelmed.

Based on the above assessment, there is a clear need for a generic simulation framework that can be used to represent the three main components of a FEWS, and which will be driven by Monte Carlo simulations of floods and their forecasts. This will allow the long-term performance of FEWSs to be assessed under a wide range of conditions. Moreover, the framework can be used as a virtual test-bed to support the design of FEWSs by exploring the sensitivity of FEWS performance to various controlling factors.

1.3 Aim and objectives

Based on the research needs outlined above, the **overall aim** of this thesis is **to build a generic framework that can simulate and identify important factors controlling the performance of a FEWS monitoring and warning a flood-prone area.** To achieve this aim, the thesis is structured around the following objectives:

1. To design a flexible generic simulation framework that can represent the forecast, warning, and response/impact/components of a FEWS.
2. To design a Monte Carlo flood and forecast generator (MCFG) applicable to a generic fluvial case and a flood-prone polder system case.
3. To design the flood warning decision component (FWDC) of a FEWS and to simulate and identify important factors controlling the flood warning reliability of a FEWS for a generic fluvial river case under deterministic and probabilistic forecast information.
4. To design the response and impact component (RIC) of a FEWS and to simulate and identify important factors controlling the economic

effectiveness of a FEWS for a generic fluvial river case under deterministic and probabilistic forecast information.

5. To apply the generic framework to a case study of the operation of a flood-prone polder system in Nanjing, China

1.4 Thesis structure

The remaining part of the thesis is structured as follows. Chapter 2 reports on a literature review covering the forecast, warning, and impact/response components of a FEWS and discusses the research needs highlighted above. This helps to introduce some key background information required to understand the rationale behind the development of the generic framework. This generic framework is explained in Chapter 3 and has three components: the Monte Carlo flood and forecast generator (MCFG), the flood warning decision component (FWDC), and the response and impact component (RIC) (Objective 1).

Chapter 4 gives a full description of the architecture of the MCFG (Objective 2). Chapter 5 integrates the MCFG and the FWDC components of the framework for simulating a generic fluvial FEWS and identifies the important factors controlling FEWS reliability under deterministic and probabilistic forecasting (Objective 3). In Chapter 6, the RIC component of the framework is added to explore important factors influencing the economic effectiveness of a (deterministic and probabilistic) FEWS based on real-time flood maps (Objective 4). In Chapter 7, the generic framework is applied to the case study, and a full description of the building of each of the components for simulating a FEWS monitoring a flood-prone polder system case is provided. This model is then used to explore factors influencing the performance of this FEWS in polder flood management operation (Objective 5). A discussion, conclusions, and recommendations for further work are finally provided in Chapter 8.

Figure 1.3 illustrates the structure of the thesis and the relationship between the objectives and chapters.

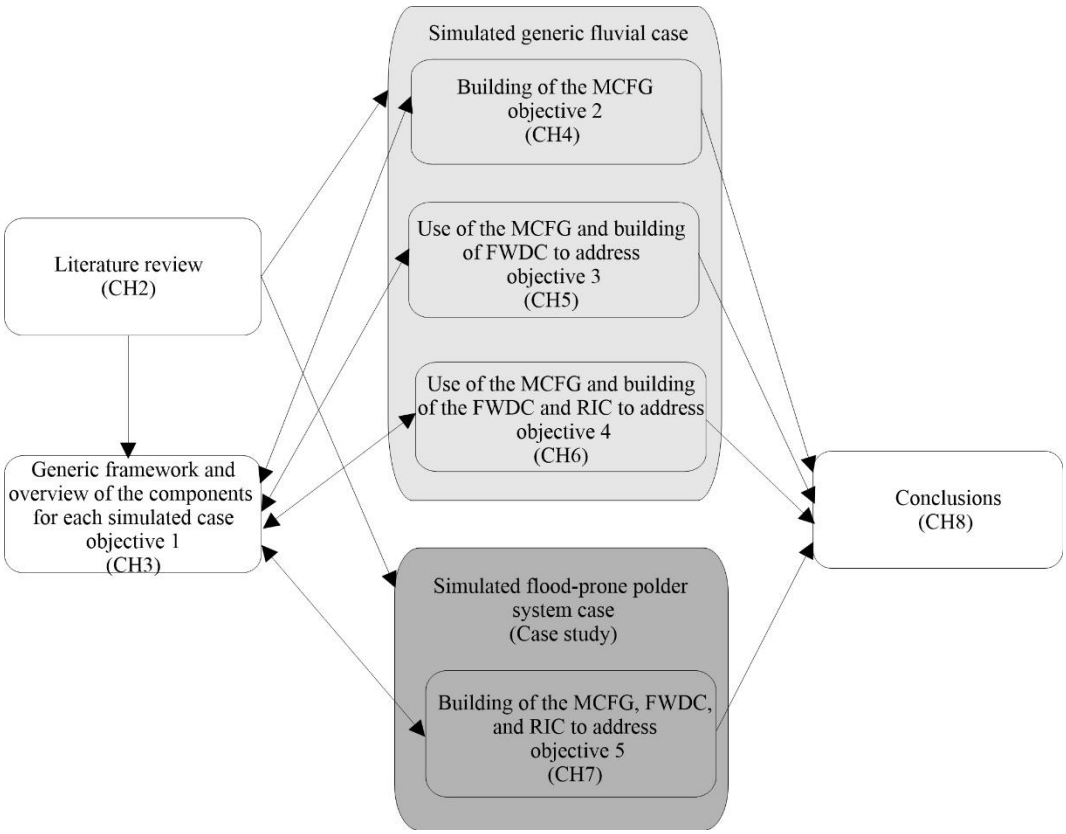


Figure 1.3: Schematic representation of thesis structure and relationship between the objectives and Chapters 2 to 8.

Chapter 2. Literature review

2.1 Introduction

Previous research on the operation, evaluation, and simulation of a flood early warning system (FEWS) is reviewed here. This chapter i) describes important factors influencing the performance of a FEWS, ii) provides some background information about the actual operation and evaluation of FEWSs, iii) references relevant studies whose aims have been to improve the performance of a single component of the FEWS or the overall performance of the system, and iv) supports the research gaps identified in Chapter 1.

The Chapter is structured as follows: In section 2.2, a full description of a FEWS is given. In section 2.3, the forecasting component of the FEWS is explained. This section describes several options available to set up a flood forecasting system (FFS) in a catchment and how they affect the quality of the forecasts. Here the trade-off between *uncertainty* and *forecasting lead time* and the types of forecast errors are described. Section 2.4 explains the methods available to estimate *predictive uncertainty* (PU) in flood forecasting and shows different uncertainty measures can be derived from them. The concept of ensembles is discussed briefly, and the section focuses more on statistical models (also known as post-processors and which are the methods considered in this research) and the concept of PU. Section 2.5 explains how flood warning decisions are made in operational FEWS, and the most common warning criteria used in a local context, and how several research works have simulated flood warning decisions considering these warning criteria. This section also describes the methods available for issuing flood warnings by using probabilistic information. Section 2.6 explains the response to flood warnings and describes several potential proactive actions that can be part of the response component of a FEWS. Section 2.7 describes some attributes used to define the performance of FEWSs. Substantial emphasis is placed on the *reliability* of flood warnings and *economic effectiveness*, and several relevant studies are referred to in this field. **The term *economic effectiveness* is used in this research to refer to the economic flood risk reduction achieved by a FEWS relative to the economic flood risk of the no warning scenario**. Section 2.8 compiles all relevant information provided in the prior sections, and other research works to describe how FEWSs are actually operated and evaluated. Section 2.9 references research

works that have analysed the performance of a FEWS through an end-to-end modelling framework. Through the development of the Chapter, important factors influencing the performance of a FEWS are described and the research gaps identified are highlighted; they are then summarised in section 2.10, and the relation between the considered topics and this research is explained.

2.2 Flood early warning in a system context

Flood warnings are part of an integrated system that can be idealized as a cascade coupling four components that are supported by several processes (Figure 2.1). That integrated end-to-end system is often referred to as a Flood Early Warning System (FEWS) (Girons Lopez *et al.*, 2017), or flood forecasting, warning, and response system (Verkade and Werner, 2011; Parker, 2017). In this research, the first designation has been chosen. The system is explained as follows.

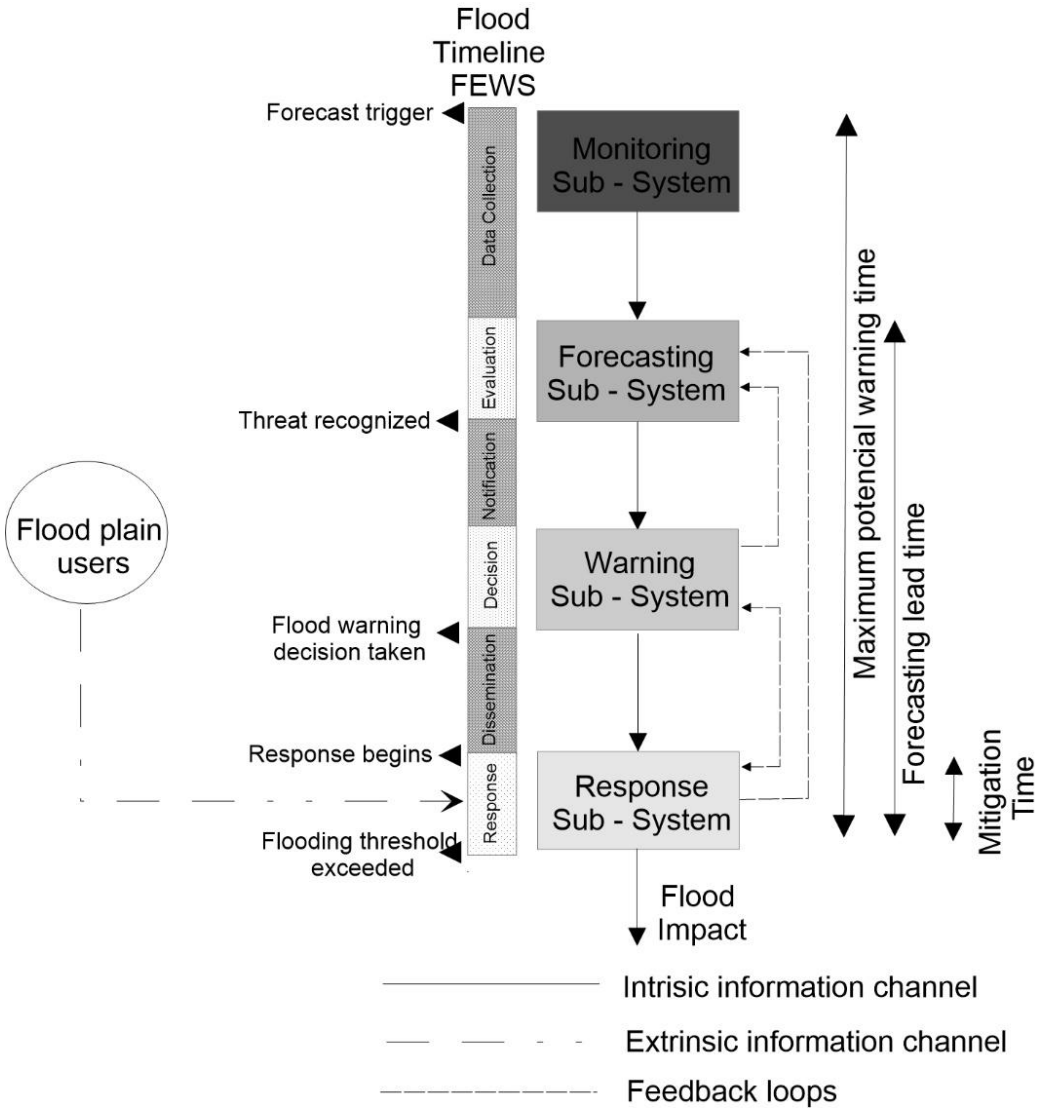


Figure 2.1: Flood early warning system
Adapted from Parker and Fordham, (1996) and Carsell *et al.*, (2004).

The first milestone in a flood event is the occurrence of a forecast trigger, and the last one is the exceedance of a stage level (flood threshold) at which flood losses, such as economic damage, trauma to residents, and potential loss of lives, start. Although trauma is important - floods can cause social and mental health problems that may continue over extended periods of time (Stanke *et al.*, 2012) - it is very difficult to quantify. As shall be seen later, the current work concentrates on economic damage only, and no attempt is made to represent psychological damage to residents, or in the worst case, loss of life.

All components of the FEWS are between these two milestones. This period is theoretically the maximum time that is available for conducting proactive action, and it is known as *the maximum potential warning time* (Carsell *et al.*, 2004). If one or more components fail within this time period, the FEWS is ineffective.

The forecast trigger controls *the maximum potential warning time*. It is a simple or compound event that is likely to precede every flood and triggers the preparation of flood forecasts. This event is defined in terms of hydrometeorological variables which are routinely monitored and/or forecasted (Krzysztofowicz, 1993). On large and medium rivers, examples of these triggers are thresholds of an observed river stage and observed rainfall intensity and duration, respectively. The monitoring may be done by a local system or may also be done by a meteorological agency that triggers the preparation of the forecasts through the issue of a severe weather warning (Parker, 2017). In any case, once the preparation of the forecasts is triggered, time is required to collect data and transmit hydrometeorological data from the field to a central site to be examined. Meteorological forecasts usually provided by meteorological agencies in the form of numerical weather predictions (NWP) may also be used in this stage. This process is called *data collection* and is carried out by technical personnel and automated detection equipment.

At the central site, forecasters use hardware, software, and hydrometeorological data and apply all their knowledge to create relevant information to recognize the potential flood threat (the forecast is generated). This process is called *evaluation* and often compares future discharge or water level with a pre-defined flooding threshold (y_T), which is often set based on experience, historical data, and/or detailed hydraulic modelling of river response. If the potential flood threat is recognized, the time known as *forecasting lead time* starts, and time is needed to

provide this forecast to the Warner (a word adopted in this research to refer to a person involved in preparing and issuing flood warnings, usually a member of a governmental entity). This process is known as *notification* and triggers *the flood warning decision* in the FEWS. In some countries, this responsibility may fall on the forecasters (where the notification time is negligible) or professionals who have the appropriate technical background to interpret the real-time forecast information (Blöschl, 2008). Nevertheless, in other cases, this responsibility rests with non-technical decision-makers (Orr and Twigger-Ross, 2009). The implementation of this decision marks the beginning of the process known as *dissemination.* The flood warnings are disseminated to local flood authorities and the population at risk and may be based on real-time flood hazard maps or pre-defined flood hazard maps obtained from a pre-flood risk assessment of the area monitored by the FEWS. This process may be carried out by media agencies, internet, text messages (SMS, MMS), door-knocking, or phone calls, whereby sometimes intermediate warning dissemination agents, such as local civil protection authorities, are involved.

Once the recipients receive the warning, the *response* begins, and the proactive actions are implemented. These actions can be done by emergency services and others, e.g., local flood defense units or military personnel, and might include, for example, raising demountable defenses or rescuing residents. The proactive actions can also be executed by the warned householders who can move and/or evacuate residential contents to a safe place (Carsell *et al.*, 2004; Rai *et al.*, 2020). The time between the response beginning and onset of the flooding is called the *mitigation time* or *warning lead time* and is the time available for mitigation. As can be noted, this time is different from the *forecasting lead time* due to the time needed to make the warning decision and disseminate warnings (Carsell et al., 2004).

Although it is not always the case, there is usually an information flow among components of the FEWS. This information can be either intrinsic, when it is part of the system and incorporate formal warnings, or extrinsic, which involves informal warnings used by floodplain users(Parker and Fordham, 1996). Therefore, the success of the FEWS heavily depends on how effective the communication is among the personnel involved in each component. In all of the above, it is implicit that the flows of information take place in real-time. Therefore, factors associated with the failure of instruments/telemetry system (FITS), the failure of

communication networks (FCN), and human errors in the forecast and warning process (HFWP) can affect the FEWS performance.

2.3 Flood forecasting systems (FFSs) and sources of uncertainty

A FFS is part of a FEWS and has an associated *forecasting lead time*, which is a key technical factor in the integrated system as it provides time in advance to the concerned authorities to warn and take effective proactive actions in the face of a potential flood. In essence, the FFS must help the concerned authorities answer as accurately as possible the following two questions: What will be the magnitude of the flood event? And when is it going to arrive? Having an exact answer to these questions would imply having a perfect forecast, which is impossible as flood forecasts are inherently uncertain. In the last two decades, a lot of research has focussed on flood forecasting techniques with the aim of reducing this uncertainty. This literature review does not attempt to provide a comprehensive description of all these studies but focuses on the flood forecasting techniques that are considered or assumed in this research, and in particular, on uncertainty (for a detailed description of all flood forecasting techniques, the reader is referred to Hapuarachchi *et al.*, (2011) and Jain *et al.*, (2018)).

Thus, through a literature review, this section of the chapter explains the concept of *uncertainty* and the type of forecast errors associated with it. This section also explains the forecast chain often used in FEWS and how the FFSs used in this chain are related to the *forecasting lead time* and *predictive uncertainty*, PU.

2.3.1 Sources of uncertainty

Uncertainty can be defined in simple terms as the *lack of knowledge of the observed value* (Gouldby and Samuels, 2009). In flood forecasting, it produces a difference between the forecast and "true" value of the quantity being forecasted, frequently referred to as the predictand. That uncertainty can be split into two types: aleatory and epistemic uncertainty. The aleatory uncertainty is associated with the natural variability and describes the inherent randomness in nature, whereas the epistemic uncertainty is associated with the limited knowledge of the system being studied and with the inability to measure and model the physical world(Wasson, 2016). These two uncertainties are included in the formal assessment of uncertainty in

flood forecasting, and only the epistemic uncertainty can be reduced (Boelee *et al.*, 2019).

2.3.2 Uncertainty and lead time

There is a trade-off between the *forecasting lead time* and *uncertainty* in flood forecasting (Blöschl, 2008; Parker, 2017). This issue will be explained through an illustrative example (Figure 2.2) that shows a community threatened by river floods due to its low standard of protection (SoP), which needs to install an FFS as a part of a FEWS to improve the population resilience to this natural hazard. In this case, it is assumed forecasts of future discharges or water levels at the forecast points (cross-sections of the river adjacent to the at-risk community).

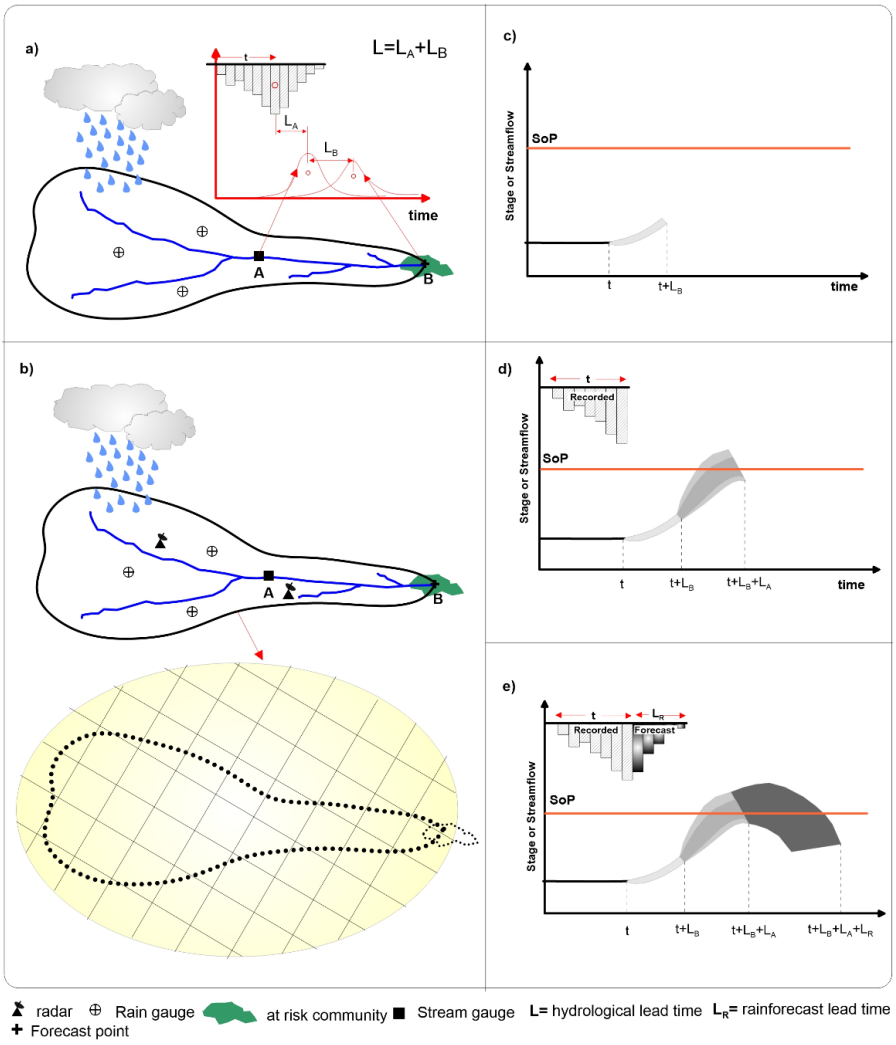


Figure 2.2: Flood forecasting techniques, forecasting lead time, and uncertainty in flood forecasting

Adapted from O'Connell (2018). An illustrative example to explain the trade-off between *forecasting lead time* and *uncertainty in flood forecasting* through the potential forecasting models described in Table 2.1. The flooding threshold y_T is defined by the standard of protection (SoP) of the at-risk community. The light and dark grey between the lead times $t + L_B$ and $t + L_B + L_A$ illustrates the uncertainty associated with gauge-based and gauge-radar based QPE, respectively.

As can be seen in Figure 2.2, the catchment area is monitored by a gauging station (point A), and by a telemetric rain gauge network in the upper area. The forecast point is represented by point B (Figure 2.2a). Due to the catchment's characteristics, the lag between points A and B is mainly dominated by the routing of the flood through the main channel. Thus, one potential flood forecasting technique to be used may be a channel routing model based on the gauging station's records at A. Assuming that there is a good quality of the hydrologic information, as well as of the predictions, which is usually easily fulfilled by this flood forecasting technique (Barbetta *et al.*, 2016), one would expect that this forecasting model would have a small level of uncertainty. However, the *forecasting lead time* of this model would be equal to the lag time between points A and B, i.e. L_B, (Figure 2.2c), which would be the smallest one compared with other flood forecasting techniques that can be applied in the catchment.

To increase the *forecasting lead time*, one can consider the telemetric rain gauge network and use, in addition to the flood routing model, a rainfall-runoff model by using gauge-based catchment rainfall estimate as an input which can be recorded until the time of forecasting t. This time is, in essence, defined by the forecast trigger that, in this example, can be a rainfall-based variable. With this forecasting chain, one can gain *forecasting lead time*; however, one also increases the *uncertainty* (Figure 2.2d). In this situation, the main source of uncertainty comes from rainfall data (Todini, 2004) as, for practical reasons, the density of rain gauges in the network will often be too low to properly characterize the spatial and temporal distribution of rainfall throughout the catchment. This problem can be partly solved using weather radars (Cluckie and Han, 2000; Reichel *et al.*, 2009) since they offer an alternative method to capture rainfall data in remote areas of the catchment and enhance the gauge-based catchment rainfall estimate (referred to sometimes as a Quantitative Precipitation Estimate (QPE) (Arheimer *et al.*, 2011; Adams, 2016). In this context, a forecasting chain based on gage-radar based QPE that includes a good bias correction technique will have the same *forecasting lead time* of the gauge-based one, i.e., $L = L_A + L_B$. However, its level of *uncertainty* is expected to be smaller (Figure 2.2b and Figure 2.2d), the residual uncertainty will be due to the remaining uncertainty in the rainfall estimate and the rainfall-runoff model error. Finally, suppose one desires to increase the *forecasting lead time* even more, i.e., longer than the hydrological lead time L (catchment response). In that case, one

must add a rain-forecast lead time by using rainfall forecast, which is typically derived from quantitative precipitation forecasts (QPFs) by NWP, i.e., from a Numerical Weather Prediction model. Nowadays, a QPF is delivered in the form of an ensemble forecast which mainly encapsulates the uncertainty deriving from the NWP initial conditions (Cloke and Pappenberger, 2009; Liguori *et al.*, 2009). Shorter lead time precipitation forecasts can also be derived in the form of Nowcasts based on the propagation of weather radar images forward in space/time using a nowcasting algorithm (Heuvelink *et al.*, 2020). A forecasting chain under these considerations records the rainfall until the time of forecasting *t* and forecasts future rainfall for a time interval into the future L_R. Even though this forecasting chain provides the longest lead time relative to the previous scenarios, the uncertainty level deriving from the Nowcasting and/or the QPF for the lead-time extension Lr is the highest due mainly to the uncertainties related to meteorological predictions (Marty *et al.*, 2013) (Figure 2.2e).

There are several other sources of uncertainty that can be substantial in the flood forecasting chain. For example, errors associated with initial conditions, e.g., soil moisture, are particularly important when using rainfall-runoff models. Table 2.1 describes the four scenarios illustrated in Figure 2.2 and their primary sources of uncertainty.

As one can see, this example explains clearly why uncertainty increases as the lead time increases and the rationality of the trade-off between these two variates. This rationality is used in this research to explore the reliability of flood warnings and the economic effectiveness of river flood warnings associated with different *forecasting lead times* in Chapters 5 and 6, respectively.

The preceding depiction of how uncertainty increases with lead time implicitly assumes that data assimilation/forecast updating is carried out as new data are received at each time point, so that, at the current time point t, the model forecast and the observed value agree within the limits of measurement error, before the next forecast is issued. The effect of data assimilation dies out as the lead time increases, and the forecast reverts to the simulation mode model forecast, i.e., that resulting from the forcing input variable(s). This is also the case with NWP model forecasts.

Table 2.1 Source of uncertainty for four potential flood forecasting techniques in a FEWS
The four forecasting models described in this table are the ones described in Figure 2.2.

Scenario	Model configuration option.	Forecasting lead time	The primary source of uncertainty	Levels of uncertainty
1	Flood routing model	L_B	• High Flow Ratings • Hydraulic/Routing Model Structure and Parameters • River channel/floodplain • Survey	The lowest one
2	Flood routing model+ rainfall-runoff model (precipitation data from rain gauge network)	$L = L_A + L_B$	• Type of rainfall event (convective, frontal, orographic, etc.) • Rainfall-Runoff Model Structure and Parameters • Antecedent Conditions	Higher than scenarios 1 and 3.
3	Flood routing model+ rainfall-runoff model (precipitation data from rain gauge network and/or weather radars)			Smaller than scenario 2.
4	Flood routing model+ rainfall-runoff model (observed rainfall and quantitative precipitation forecast (QPF) from a numerical weather prediction (NWP) model)	$L_R + L$	• Type of rainfall event (convective, frontal, orographic, etc.) • Nowcasting Algorithm Structure and Parameterization • NWP Model Structure and Parameterization • Rainfall-Runoff Model Strucure and Parameters	The highest one.

Also, it is important to note that the example shown in Figure 2.2 assumes that forecasts are done by using a rainfall-runoff model and/or flood routing model that can be physically based or conceptual. Note, however, that forecasts can also be generated through data-driven models that rely on historical data without directly

considering the underlying physical hydrological processes. These methods can represent the existing nonlinear rainfall-runoff relationship with high computational speed. The sources of uncertainty of these models are related to the aleatory uncertainty - included in the training data - and the inability of the algorithm to represent the observed values. The popularity of advanced data-driven models, e.g., machine learning models, has dramatically increased in recent years (Mosavi et al., 2018). Machine learning methods, e.g., fuzzy systems or artificial neural networks, have been used for forecasting rainfall, discharge, water levels, or for assimilation/forecast updating (Tareghian and Kashefipour, 2007; Kambalimath and Deka, 2020).

2.3.3 Types of forecasting errors

Given the existence of inherent uncertainty, flood forecasts are subject to several types of forecasting errors, which have to do with the lack of fit between the predictand, often a discharge or water level, and predictor. Based on a forecast and observed hydrograph of a potential flood, they are classified as amplitude, phase, and shape errors (WMO, 2011), and their presence also affects the forecast of the main variables involved during the forecast process, i.e., peak (y), timing (t_y), and volume (V) of the potential flood.

The amplitude error implies both errors in the magnitude (e_y) and volume (e_V) (Figure 2.3a). They often occur due to structural deficiencies in the hydrological model, parameter estimation errors, and errors in the input and/or output data. The phase errors show errors in the timing of the hydrograph (e_{t_y}), i.e., although y is correctly estimated, it is delayed or advanced in time (Figure 2.3b). Finally, shape errors may show e_y, e_V, and e_{t_y}. This type of error may be attributable to the routing component of the model not representing well the redistribution of the generated runoff volumes over time (Figure 2.3c). If the error is systematic for any of these scenarios, it is said that the forecasts are biased. FFSs operating in non-updating mode, i.e., the ones that do not consider real-time information to modify forecasts, are more likely to have these errors than the FFSs operating in updating mode, i.e., the ones that consider forecast updating based on recent hydro-climatic observations (this is a post-processing technique named data assimilation). The literature offers several bias-correction methods when dealing with streamflow

forecasts (Hashino *et al.*, 2007). These methods often address the issue of the bias in the mean value of the predictand, and the issue of the bias in the variance of the predictand has been poorly analysed in flood forecasting and warning process.

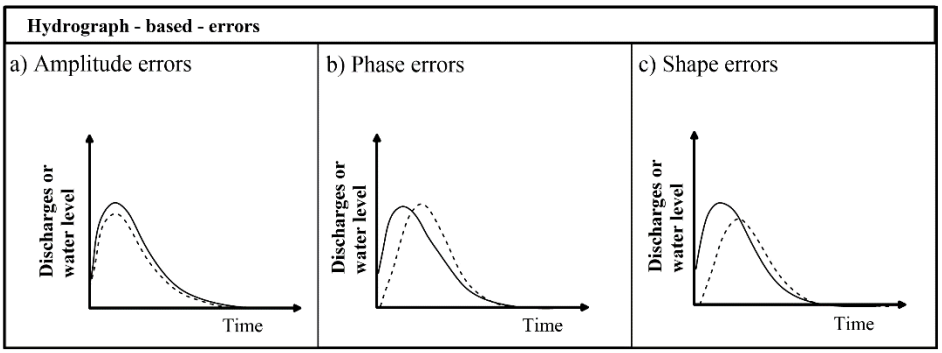


Figure 2.3: Types of forecast errors
Adapted from WMO (2011).

In this research, a potential flood is represented by its peak flow ordinate, i.e., y, and only the error e_y is considered in the representation of its forecast. The primary impact is created by the magnitude of the flood peak. The timing of the flood can be considered a secondary impact and neglecting e_{t_y} mainly affects the *economic effectiveness* of flood warnings because the damage saved by a proactive action depends on the *mitigation time,* which, in turn, depends on the onset of the flooding. The error e_{t_y} is not considered in the analysis done in Chapter 6 since it is assumed that e_y will dominate over these timing errors. The impact of the bias in the mean and variance on the reliability of flood warnings is explored in Chapter 5.

2.4 Estimation of predictive uncertainty (PU)

From more than one decades ago, the concept of uncertainty in flood forecasting has started to be considered more and more in research and practice (Krzysztofowicz and Kelly, 2000; Todini, 2004; Cloke and Pappenberger, 2009). This issue has been studied through two philosophies: ensembles methods and statistical methods (also known as post-processors)(Boelee *et al.*, 2019). Ensembles methods define *uncertainty* from a set of plausible forecasts, whereas the statistical methods use prior forecast and observed values to define PU through statistical analysis. Some methods also use these two philosophies to estimate it.

Ensembles methods assume that the space of the forcing data, forecasting model structure, and parameters can be defined. Thus, ensembles are often obtained by running the forecasting model several times, frequently from different initial

conditions, and where each run is within this assumed space (Xuan *et al.*, 2009). Since ensemble methods measure the spread of the forecasts descending from these multiple scenarios, it is said that they do not provide an appropriate measure of PU, but rather a measure of forecasting sensitivity (Todini, 2017) (Figure 2.4a).

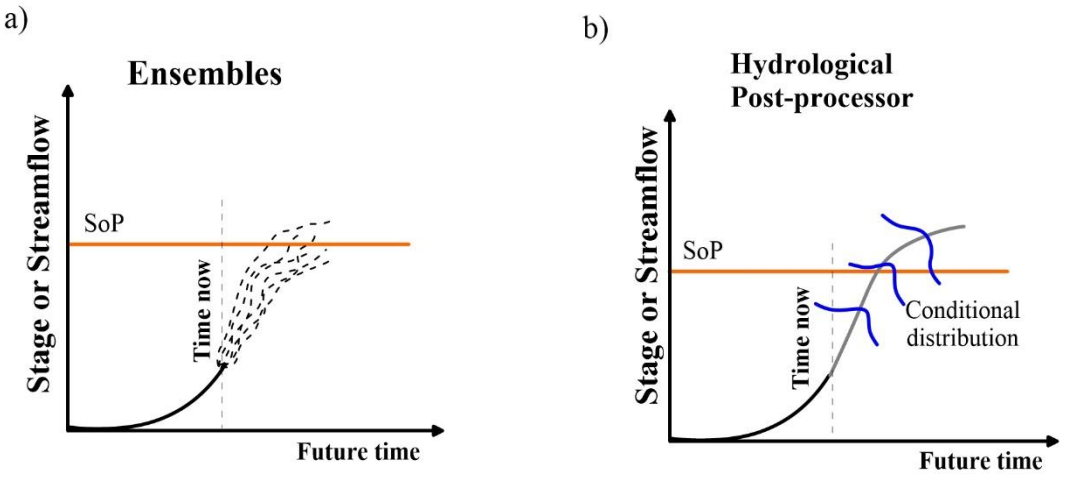


Figure 2.4: Illustration of methods used to characterise (a) uncertainty in forecasts and (b) predictive uncertainty

The flood even is defined by the Standard of protection (SoP). a) Uncertainty in forecasts described by ensembles and b) predictive uncertainty defined by conditional distributions for different forecast horizons.

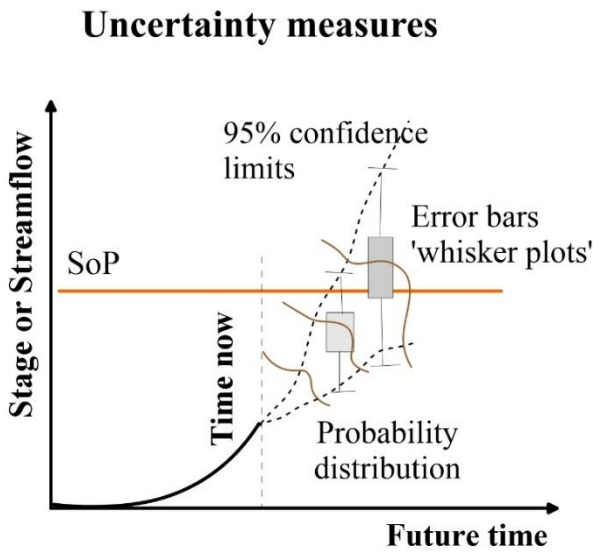


Figure 2.5: Uncertainty measures

From the methods used to quantify uncertainty from ensembles, several uncertainty measures (confidence limits, error bars, probability distributions) can be obtained. Note that in practice, only 1 or 2 measures would be used on the same figure. The flood event is defined by the standard of protection SoP.

Statistical methods assess the uncertainty in flood forecasts in terms of PU. They use historical observed and predicted values to derive PU as the predictive density of the unknown future quantity conditional on the forecasted one. PU is quantified when forecasting a predictand (often water level or discharges) with a specific forecast horizon (*forecasting lead time*) (Figure 2.4b).

Figure 2.5 illustrates the uncertainty measures that can be derived from the ensembles. Given that this research considers the statistical philosophy to represent PU, the PU concept, and the methods used to estimate it, are analysed in the following subsections.

2.4.1 Predictive uncertainty (PU) definition

In studies of hydrological forecasting, the terminology "*predictive uncertainty*" has been widely used. PU is defined by Todini (2016) as the probability that the predictand, typically the observed peak water level or discharge at a specific cross section at the time $t + L$ (where L is *forecast lead time*) conditional on a single or multiple predictor(s), namely model prediction(s). The predictand is conditional on the structure, parameters, and forcing data of the forecasting model(s). The terminology "*predictive uncertainty*" has been popular because it highlights that it is associated with the uncertainty around the prediction of the predictand, rather than the "*simulation uncertainty*" (SU), also known as "emulation uncertainty" or "validation uncertainty", that defines the uncertainty of the predictor given the knowledge of what actually occurred. SU, in essence, defines the skill of the forecasting model to reproduce the reality (Klein *et al.*, 2016). If one assumes peak water level as the predictand, i.e., y, and only one predictor (the forecast peak water level of a forecasting model), i.e., $\hat{y}$, the joint probability of y and $\hat{y}$, PU and SU are defined by $f(y, \hat{y})$, $f(y_t|\hat{y}_t)$ and $f(\hat{y}_t|y_t)$, respectively. Since, in the case of SU, $\hat{y}$ is the uncertainty quantity, it can only be based on past observations and not in predictive mode, whereas PU can be used in both simulation (or hindcast) and predictive mode. Figure 2.6 illustrates the concepts of SU and PU.

It is important to realise that PU encapsulates all the different sources of uncertainty in the forecast chain into one measure, thus making it unnecessary to consider the sources themselves.

Throughout this thesis, the terms 'predictive uncertainty' and 'forecast uncertainty' are used interchangeably, where it is understood that the latter term has exactly the same interpretation as the former.

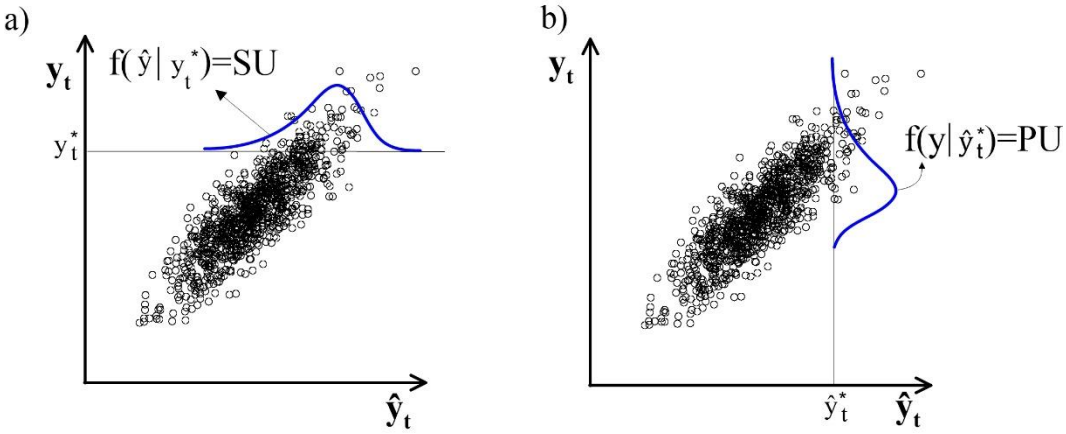


Figure 2.6: Illustration of SU and PU

a) Simulation uncertainty (SU). Once the joint distribution between the observed and forecast values is built, knowing an observed value y_t^* , one can evaluate the uncertainty of the predictor $\hat{y}$ (forecasting model). b) Predictive uncertainty (PU). With the joint distribution, knowing the predictor $\hat{y}_t^*$, one can assess the uncertainty on the predictand. Adapted from Todini (2016).

2.4.2 Predictive uncertainty (PU) estimation

The statistical methods (which are the methods considered in this research), in essence, calculate the forecasting model's error and assume the model's errors, obtained from past observations and their forecasts, are representative of the uncertainty in the future. These methods estimate the PU of a predictand (often water level/discharge) for a specific forecast horizon (*forecasting lead time*) and location. Thus, statistical methods have been applied to a fixed location with discharge/water level data (Boelee *et al.*, 2019).

The statistical methods most commonly used in hydrology are the Hydrological Uncertainty processor (HUP)(Krzysztofowicz and Kelly, 2000), Model Conditional Processor (MCP)(Todini, 2008; Barbetta *et al.*, 2016), and Bayesian Joint Probability (BJP) model (Wang *et al.*, 2009; Zhao *et al.*, 2015). All of these methods, which have been reviewed by Li *et al.* (2017), share similarities; for example, i) they use parametric approaches based on Bayes' theorem to derive the distributions needed to estimate PU; ii) a sample of forecasts and the resulting observations, obtained from historical records or simulated forecasts (hindcast), is used to identify the distributions involved in the Bayes' theorem; iii) the distributions of the original samples are transformed to Gaussian to apply the Bayesian inference technique; and iv) the samples used are comprised of values that represent the four different states when dealing with flood forecasting, i.e., peak flow, base flow, and transitory states occurring during the rising and recession limbs.

In the same way, these methods have their own particularities. For example, HUP takes into account the strong autocorrelation property of hydrological time series to base the derived probability distributions on recent past observations (Krzysztofowicz, 1999a; Bogner and Pappenberger, 2011). MCP is an alternative to HUP, but it has the advantage that the derivation for univariate situations (a single model provides probabilistic forecasts at a specific *forecasting lead time* and location) can be conveniently extended to multivariate situations, which can be used, for example for multimodel, multisite, and multi-lead time problems (Coccia and Todini, 2011). Both HUP and MCP use the normal quantile transformation (NQT) as a Gaussian-transform method (Bogner *et al.,* 2012). BJP differs from these two models in that: i) the log-sinh transformation is used instead of NQT; and ii) it considers the uncertainty of the parameters used to describe the PU, where the Bayesian inference technique is applied for the estimation of these parameters (Zhao *et al.*, 2015). MCP and BJP deal with the exiting heteroscedasticity of the residuals when forecasting streamflow – that is, the error between forecasts and observations tend to be much higher at high flow than at low flow- through the use of the multivariate truncated Normal (MTN) distributions and the log-sinh transformation, respectively.

The main disadvantage of these statistical methods is that they rely on observed values. This problem is particularly important for the methods that use NQT as a Gaussian-transform method. It might happen that, during operational use for forecasting, the sampled data points in the normal space fall outside the range of the historical sample, and the inversion of the empirical NQT is not possible. These methods deal with this issue as extreme values are rarely observed in the historical records due to the limited amount of available data. Thus, to estimate future possible extreme values, different approaches have been proposed, which in essence seek to identify the best curve for fitting the sample distribution tails. One of them is, for example, to extend the historical observed sample with extreme values estimated by fitting a Peaks-Over Threshold (POT) model to the upper tail of the sample and apply a non-parametric regression method called Generalized Additive Model (GAM) as an extrapolation method (Bogner *et al.*, 2012). Other approaches suggest extending the sample by the hyperbolic approximation for the uppermost-tail of the distribution (Seo et al., 2006) or applying a linear extrapolation on a number of points in the tails of the sample distribution (Weerts *et al.,* 2011). Coccia (2011)

suggests applying a specific model to the tails of the historical sample by setting, for the upper tail, the maximum value, for which the probability is assumed to be equal to 1, to twice the maximum value ever observed, and, for the lower tail, by setting the minimum value, for which the probability is assumed to be null, to zero.

This research does not explicitly use a statistical method to estimate PU from observed data. Rather, it selects realistic parameter values for the distributions used in deriving PU (Figure 2.6b) within a Monte Carlo sampling framework; these parameter values are supported by the analysis of observed flood peak data. A joint distribution is assumed for the predictand and predictor. Then, using the method known as the conditional approach (Lewis and Orav, 2018), bivariate values associated with a conditional distribution (representing PU) are generated from this assumed joint distribution. This is done for a simulated fluvial case (Chapter 5 and Chapter 6) and the case study (Chapter 7). Chapter 3 describes the predictands considered in each case, whereas Chapter 4 details the statistical concepts used in this approach.

2.5 Warning decisions in flood early warning systems

The prediction problem in FEWS has been addressed by a plethora of research studies, ranging from simple to complex methods and forecasting models. In contrast, the warning problem, particularly the warning decision issue, has received relatively little attention. FEWSs often have an associated warning criterion that determines which warning decision process must be conducted. This warning criterion, in turn, depends on the flood forecasting technique used in the system. Based on a literature review, this section of the Chapter explains how flood warning decisions are actually made in operational FEWSs and the forecast type often used. Furthermore, it describes the most common warning criteria used in a local context and how several research works have simulated flood warning decisions considering these warning criteria.

2.5.1 Common forecast type and warning decision

Most FEWSs use deterministic forecasts in the warning-decision-making process. However, due to the significant progress in the last two decades in assessing and estimating PU, the probabilistic forecast is starting to be considered in operational FEWSs. For example, the Environmental Agency (EA), the entity responsible for

managing the risk of flooding in England, is transitioning toward using this forecast type through the use of ensembles (Arnal *et al.*, 2020).

Regardless of the type of forecast used, the warning-decision-making process is complex. Before issuing a warning, the Warner often meticulously assesses the forecasts and uses his/her previous experience (built up over past events) to make the best decision (Ramos *et al.*, 2010; Verkade and Werner, 2011). Other factors such as the type of event (e.g., local small flood events vs large-scale extreme flood events), the cost-benefit analysis of warning vs not warning, risk attitude (e.g., risk aversion vs risk loving), and the cultural environment in which decisions are made, also influence the warning decision (Arnal *et al.*, 2020). All these factors associated with the human behaviour of the Warner have been encapsulated in this research in a factor called the human component of the flood warning decisions (HFWD), which, as was mentioned above, affects the flood warning decision and, therefore, the FEWS performance.

2.5.2 Warning criteria in a local context

FEWSs may use for the warning criterion a single-discharge water level threshold (e.g., defining defence overtopping or first property flooding), where forecast values (often discharge or water levels) are compared with the pre-defined threshold, which is, in essence, the flooding threshold y_T. If the forecast values cross y_T, a warning should be issued (Figure 2.7a). These FEWSs do not target warnings to individual properties or small areas within the at-risk community but mainly aim to alert flood authorities of a potential flood that, in turn, uses proactive actions to mitigate the flood impact. A common proactive action in this type of FEWSs is raising demountable defenses, which can need a 24-h ahead warning to be erected. Flooding-threshold-based warning decisions have been the basis of some important research (Verkade and Werner, 2011; Dale *et al.*, 2014).

FEWSs may also use several warning criteria based on different discharge/water level thresholds. These thresholds are often associated with different return periods and are called flood warning thresholds. They split the level of impact in the floodplain into flood zones (also called flood warning areas) which allows warnings to be targeted on small areas (Sene, 2008). This warning criterion represents a more phased approach to warning or evacuating properties than the single flooding-

threshold-based criterion since each flood zone has its own flood warning threshold. In this case, a flood zone is warned if its corresponding flood warning threshold is exceeded (Figure 2.7b). These thresholds are often set based on prior hydraulic modelling and flood risk assessment of the area monitored by the FEWSs.

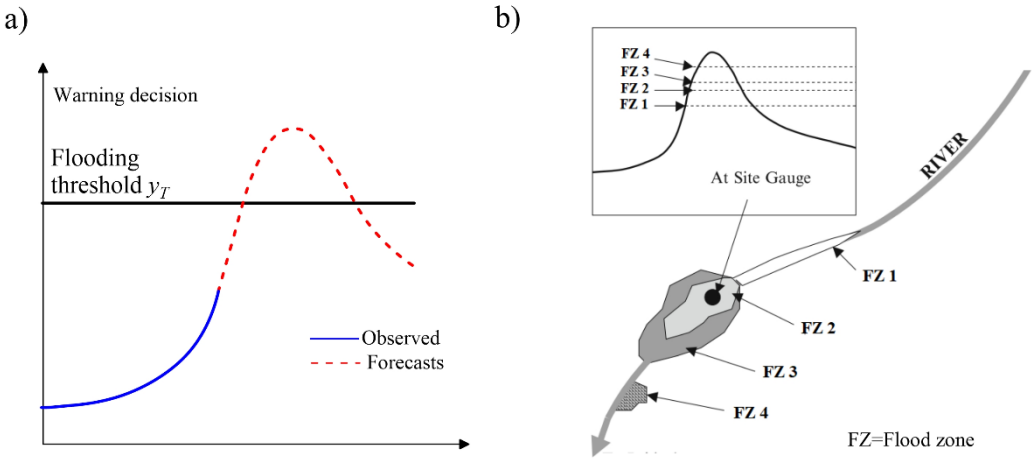


Figure 2.7: Illustration of warning criteria based on discharge or water levels thresholds
a) Single flood warning-threshold-based criterion, b) multiple flood-warning-thresholds based criterion (adapted from (Sene, 2008)).

Some FEWSs may use a hydraulic inundation/hydrodynamic model in the forecast chain to transform the hydrograph of a potential flood into the expected socio-economic impacts through real-time flood hazard maps. In this case, these maps and the flooding threshold y_T are used in the warning criterion. Warning decisions are based on the direct comparison of future discharge or water levels and the flooding threshold y_T, and the real-time flood hazard maps are used to identify individual vulnerable properties and disseminate flood warnings (Figure 2.8a). Flood hazard maps are built by intersecting the maps of the property locations and the inundation forecasting. This flood warning service is considered the most sophisticated one and is considered complex and computationally expensive, and its use has been limited in FEWSs (Fernández-Nóvoa *et al.*, 2020). Despite that, with the development of high-performance computing techniques, it is now feasible to use these flood forecasting techniques in flood warning services (Ming *et al.*, 2020; Ritter *et al.*, 2020). In some cases, a library of pre-generated hazard maps corresponding to different flood levels is created and stored, and the maps are then interpolated in real-time, which is computationally light.

Some FEWSs may use a warning criterion based on relatively simple forecasting tools, such as the rainfall threshold (RT) method (Georgakakos, 2006; Golian *et al.*, 2015). These methods allow the forecaster to use local precipitation and predict

flooding without running complex forecasting models. Furthermore, they can be used as a backup approach when there is failure or instability in the main FFS (O'Connell, 2005). In this method, a critical RT value is defined as the minimum cumulative rainfall volume necessary to cause a critical river discharge at the forecast point (a cross section of the river). A rainfall threshold curve is then built by plotting critical RT values versus time, all corresponding to the same critical river discharge. When the cumulative rainfall volume of an event crosses the RT curve, the peak discharge at the forecast point is expected to be equal to or greater than the RT curve's critical river discharge. Thus, FEWSs that use this flood forecasting technique use a warning criterion based on future precipitation (often expressed in terms QPF) and the RT curve. In that case, a warning is issued if the QPF crosses the RT curve (Martina *et al.*, 2006)(Figure 2.8b). Like the FEWSs based on a flooding threshold (Figure 2.7a), these systems do not target warnings to small areas or individual properties but aim to alert the general public and lead local flood authorities of a potential flood in the floodplain. These methods often assume a pre-defined rainfall pattern, pre-defined soil conditions, and fixed catchment characteristics, and, therefore, relatively large uncertainty in the flood predictions may come from these assumptions (Martina et al., 2006; Hapuarachchi et al., 2011; Wu et al., 2015).

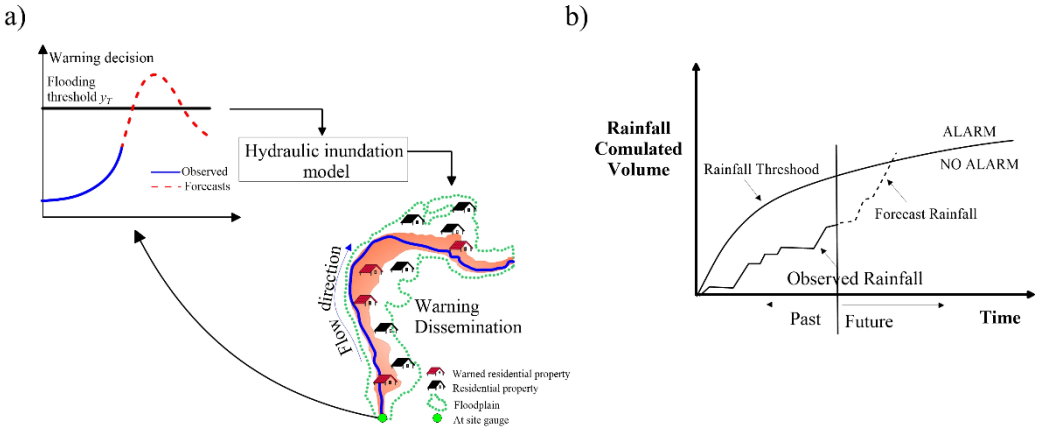


Figure 2.8: Illustration of warning criteria based on a) inundation forecasting and b) rainfall threshold curve

This research in Chapter 5 explores the reliability of flood warnings of FEWS whose warning criterion is based on real-time flood hazard maps and/or the flooding threshold y_T (Figure 2.8a and Figure 2.7a, respectively). Furthermore, in Chapter 6, the *reliability* of flood warnings and the *economic effectiveness* of a FEWS based on the first warning criterion are explored. In Chapter 7, the case study, a rainfall-threshold-based warning criterion is used. This research does not

simulate the warning criterion illustrated in Figure 2.8b; however, it takes the concept of the RT curve to adapt it to a flood-prone polder system.

2.5.3 Simulation of warning decisions

Most research works that have considered the warning decision problem in FEWSs have assumed that they can be made following a set of forecast-driven rules. For example, a common method when simulating deterministic-forecast-based warning decisions is to assume that a warning is automatically issued when the forecasting model's deterministic output(s) cross a pre-defined threshold/s. Under this assumption, Verkade and Werner (2011) show, using a quasi-analytical framework with some limiting assumptions, how flooding-threshold-based warning decisions based on deterministic-single-value forecasts, in which forecast errors are not acknowledged, lead to sub-optimal decisions in a FEWS.

The use of probabilistic forecasts allows, on the other hand, optimal flood warnings to be issued as PU is explicitly quantified. It can be used within appropriate decision-making procedures to minimize costs, economic losses, loss of life, and social disruption. These procedures can be split into probabilistic thresholds-based methods and risk-based methods. In essence, the former ones assume that a warning is automatically issued when the probability of exceedance of a pre-defined threshold of the predictand (discharge, water level, rainfall amount, etc.) exceeds a predefined probabilistic threshold. This method has been considered by assuming a warning criterion based on a flooding threshold (Verkade and Werner, 2011) or flood warning thresholds (Alfieri *et al.*, 2012; Bischiniotis *et al.*, 2019)(Figure 2.7). Probabilistic thresholds are often set in such a way that they meet specific requirements, such as maximizing the *reliability* or *economic effectiveness* of the FEWS.

Risk-based methods are often used in a Bayesian decision scheme where the predictive density, i.e., the PU, and a utility/loss function, are used to set rules based on expected values (Economou *et al.*, 2016; Todini, 2017). The utility/loss function gives information about the utilities/losses in the FEWS with and without the warning action. This function, along with the predictive density, is used in real-time. Bayesian warning decisions assume that the warning is automatically issued when the warning action's expected utility /losses are higher/lower than that without

the warning action. Bayesian FEWSs have been designed to predict floods based on rainfall (Martina *et al.*, 2006; Economou *et al.*, 2016) or water levels (Krzysztofowicz, 1993). Bayesian warning decisions are rarely used in operational systems, and they have mainly been considered in research works to show the advantage of using probabilistic forecasts in FEWS.

This research simulates warning decisions through forecast-driven decision rules. Deterministic-warning decisions assume that the warning is automatically issued when the deterministic forecast crosses a predefined threshold. The probabilistic-threshold approach is, on the other hand, used to represent the probabilistic warning decisions. These approaches are used to simulate warning decisions in a FEWS operating for a river-flood-prone area (Chapters 5 and 6) and a flood-prone polder system (Chapter 7).

2.6 Response to flood warnings

The aim of a FEWS is to reduce human and damage losses through the implementation of several proactive actions triggered by the flood warnings. These actions represent the response of the FEWS and define the benefits of the system. Priest *et al.*, (2011) developed a model to quantify the benefits of a FEWS based on several response pathways to flood warnings or risk reduction actions (RRAs). In their work, eight principal RRAs were identified, which cover the mitigation of the economic and human losses (Figure 2.9). These RRAs provide a good insight into the potential proactive actions that can be part of the response component of a FEWS and are explained as follows.

Six of these RRAs relate to damage-reducing responses and two to the reduction of human losses. The former responses cover a flood defence operation (FDO) and community-based options (CBO) that consider, for example, community pumping schemes or the placement of measures to protect properties at a community scale. The damage-reducing responses also involve actions at the household level, such as contingent resilience measures (CRM) or the contents moved or evacuated (CME) by householders. CRM is a set of planned measures to be conducted in advance of a flood, such as the use of sandbags. Watercourse capacity maintenance (WCM) also reduces flood damage. Activities to mitigate the economic damage in business, such as moving equipment or stock out of the path of floodwater, are also included

in the damage-reducing responses. On the other hand, the RRAs associated with the reduction of human losses relate to search and rescue (SAR) and evacuation measures (EVAC).

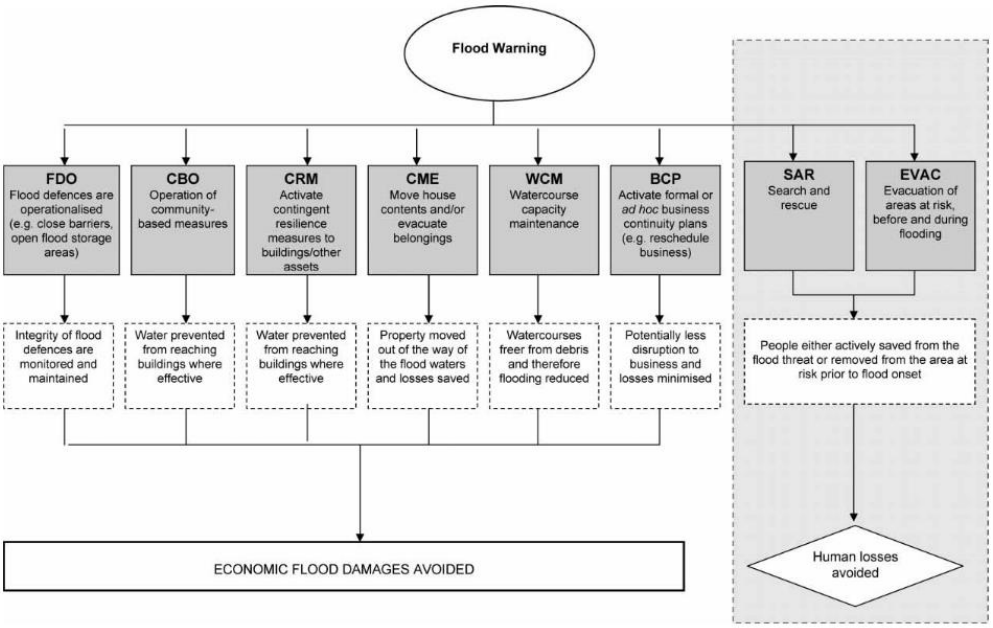


Figure 2.9: Potential RRAs used in a FEWS
Source: Priest *et al.*, (2011).

Several research works that have studied the response component of the FEWS through a modelling framework have considered one of these RRAs. For example, Dawson *et al.* (2011) and Liu and Lim (2018) used an agent-based modelling framework to explore the benefits of EVAC in a FEWS. Verkade and Werner (2011) and Bischiniotis *et al.* (2019) used a quasi-analytical framework to analyse the *economic effectiveness* of a FEWS based on CME and CMR, respectively. In this research, the response component of the FEWS is represented through CME for the fluvial case (Chapter 6) and CRM for the flood-prone polder system case (Chapter 7). The CME is simulated by using the functions proposed by Carsell *et al.* (2004), which represent the damage reduction of this RRA in a floodplain property as a function of the *mitigation time* (which is assumed in this thesis to be equal to the *forecasting lead time*) and flood depth. The functions were developed from the statistical analysis of surveys completed by floodplain management and flood damage experts. The CRM in the case study represents a pumping scheme operating in the flood-prone polder before or during storm events to remove water from the inner rivers to the outer rivers, to enable water to drain from the polder areas into the inner rivers. An overview of these RRAs is provided in Chapter 3,

and a detailed explanation of its simulation is provided in each corresponding Chapter.

As was mentioned at the beginning of this section, the RRAs shown in Figure 2.9 define the benefits of a FEWS. Particularly, the RRAs relate to damage-reducing responses define the *economic effectiveness* of the system, which is considered as one of the performance measures of a FEWS (Parker, 2017). Since this performance measure is one of the measures to be explored in this research, the potential damage saved associated with damage-reducing responses are analysed in the next section.

2.7 Evaluation of the performance of a FEWS

Parker (2017) suggests several measures to define the performance of a FEWS. He split them into technical and social measures. The former ones consider technical aspects associated with forecast quality, such as the accuracy, reliability, timeliness of flood warnings and flood detection, as well as aspects related to the geographical coverage. These measures can be characteristics of the FEWS or be evaluated based on a record of observed data and their forecasts. On the other hand, social measures are evaluated based on social survey responses and measure the acceptance of the flood warning service. Social aspects include information quality and satisfaction of the warning service, life protection, and economic factors such as damage reduction and benefit-cost ratio. Social measures could also consider the reduction of psychosocial impacts of flood events, as it has been proved that floods can cause social and mental health problems that may continue over extended periods of time (Stanke *et al.*, 2012); however, these measures are rarely considered. From these measures, *reliability* and *economic effectiveness* have been the attributes of a FEWS that have received more attention in research works. This research is based on these two performance attributes, and, therefore, substantial emphasis is placed on them in this section. This literature review supports the research gaps identified in Chapter 1, which are, in turn, highlighted in Chapters 5, 6, and 7.

2.7.1 Reliability

The *reliability* of the flood warnings is one of the foremost attributes of a FEWS. In general terms, *reliability* is defined as an object's capability to achieve a required function under stated conditions for a stated period of time (George and Modarres, 1994). Thus, Sättele *et al.* (2015) state that reliable warning systems for natural

hazards can be considered as those that detect all hazard events in a timely manner and transfer the warning to the affected people, leading to actions that avoid damage and loss of life. They also split the reliability of a warning system for natural hazards into technical reliability and inherent system reliability based on this concept. The former has to do with the system's ability to work correctly in extreme circumstances, and the latter with the skill of the system to detect and alert the hazard event.

In terms of flood warning systems, the *reliability* of the flood warnings is often evaluated only in terms of the *inherent system reliability*. Therefore, it is defined by skill scores that determine warning decision-observations combination results. Thus, it has been analysed based on the concept of signal detection theory through the probability of detection (POD) and the probability of false detection, also known as false alarm rate (F)(Krzysztofowicz *et al.*, 1994). It has also been evaluated in terms of hits, missed events, and false alarms for different thresholds (Montesarchio *et al.*, 2009; Alfieri *et al.*, 2012). In line with this, Parker (2017) defines POD, F, and the false alarm ratio (FAR) as reliability measures of a FEWS (Table 2.2). Other works define this attribute of the FEWS through the critical success index (CSI), also known as threat score (Jolliffe and Stephenson, 2012).

The works mentioned above have evaluated the *reliability* of flood warnings based on whether a warning preceded a flood event's occurrence or non-occurrence in the floodplain *or* in a specific flood risk zone, where a pre-defined threshold defines the occurrence of the event (Figure 2.7). The skill scores used to define this attribute, in essence, gives a "snap-shot" of the *reliability* of the flood warnings in the floodplain *or* a specific flood risk zone. Even though these works suggest methods to reasonably estimate the *reliability* of flood warnings, they do not consider the uncertainty in the estimation of the flood magnitude. Note that, due to the forecast's inherent uncertainty, there is always a difference between the magnitude of the forecast and observed flooding, and, therefore, a difference between the warned and flood properties. This undoubtedly impacts the *reliability* of flood warnings. For example, Parker *et al.* (2007) found in a flood-warning-customer-oriented research project that, due to a bad performance in the dissemination procedure, only 37.5% of flooded properties received a warning before flooding. This occurs due to not only poor communication but also due to uncertainty in the estimation of the flood magnitude.

Therefore, a more reliable approach would be the one that evaluates the *reliability* of flood warnings based on whether a warned property was or not subsequently flooded. This target information for this approach is the number of warned and flooded houses after the observed realization of a potential flood in the area monitored by the FEWS. To the best of the author's knowledge, a modelling-based framework that explores the *reliability* of flood warnings based on this criterion has not been reported in the literature.

This criterion, which is becoming increasingly relevant to operational FEWS as flood hazard maps are being used to assess which properties might be flooded (Figure 2.8a), is called in this research a floodplain property-based criterion (FPC) and is used to explore flood warning reliability for a simulated fluvial case in Chapters 5 and 6. In these Chapters, the FPC is compared with the criterion that evaluates the *reliability* of flood warnings based on whether a warning preceded a flood event's occurrence or non-occurrence in the floodplain, where the flooding threshold y_T at a specific river cross-section defines the occurrence of the event. This criterion is called in this research a flooding threshold-based criterion (FTC) and can be used to evaluate the reliability of flood warnings of FEWS with the warning criterion illustrated in Figure 2.7a. The same skill scores can define the *reliability* of flood warnings in the FTC and FPC. Table 2.2 indicates the interpretation of the skills scores POD and FAR in the FTC and FPC.

Table 2.2 Metrics to evaluate the flood warning performance.

Metric	Abbreviation	In the FTC, answer the question:	In the FPC, answer the question:	Equation	Interpretation
Probability of detection or hit rate	POD	What is the probability of an observed event being warned in the floodplain?	What is the probability of a flooded house being warned in the floodplain?	$POD = \frac{h}{h+m}$	Ranges from 0-1. Perfect forecast 1.
False alarm ratio	FAR	What is the probability of a forecast event being incorrectly warned in the floodplain?	What is the probability of a warned house being incorrectly warned in the floodplain?	$FAR = \frac{f}{h+f}$	Ranges from 0-1. Perfect forecast 0.
Legend: h=hits; m=misses, f=false alarms.					

2.7.2 Economic effectiveness

The term *economic effectiveness* has been used to refer to the flood risk reduction of a FEWS relative to the flood risk of the no warning scenario (Sättele *et al.*, 2015; Parker, 2017). As was mentioned in section 2.6, the RRAs relate to damage-reducing responses define the *economic effectiveness* of a FEWS which is, in turn, considered as a performance measure of the system. The potential damage saved associated with these RRAs has been often estimated through social survey research projects. From the RRAs shown in Figure 2.9, CME was perhaps the first RRA that was researched. Day (1970) was the first to do this. Based on research conducted in Susquehanna River basin, he developed what is commonly known as the 'Day curve', which shows the relationship between the mitigation time and potential damage saved (expressed in percentage) associated with CME. Based on this curve, Day (1970) suggest predicting the *economic effectiveness* of the FEWS as a percentage of a metric that defines the flood damage of the floodplain without the warning service, where the expected annual damage (EAD) is often used for these purposes. After this work, a long series of social survey research projects were developed to improve the Day's method. For example, Carsell *et al.* (2004) criticise the Day's method and mentions that the flood depth is an important factor influencing the damage reduction. Therefore, it should be considered in the estimation of the potential damage saved. They also mention that Day's method is too optimistic as it assumes that floodplain residents act rationally and efficiently. In this context, based on a social survey research project, Carsell *et al.* (2004) developed damage reduction curves associated with CME as a function of the mitigation time *and* flood depth and suggest a method to estimate the *economic effectiveness* of a FEWS as the difference between the EAD of the floodplain associated with and without a warning scenario. The latter EAD is estimated based on these damage reduction curves. The *economic effectiveness* is then reduced by considering an efficiency parameter of the RRA.

The potential damage saved by other damage-reducing responses has also been studied. For example, Priest *et al.* (2011) propose a method to estimate the economic effectiveness of a FEWS considering several RRAs in the response component of the system. Their method considers all the damage-reducing responses shown in Figure 2.9 and defines the *economic effectiveness* of the FEWS as:

$$FDA = [EAD\ x\ PFP\ x\ FDO] + [EAD\ x\ UFP\ x\ (CBO + WCM + BCP + CRM + CME)] \quad \text{Eq. 2.1}$$

Where FDA is the potential flood damage avoided by the FEWS (*economic effectiveness*), PFP and UFP is the proportion of properties at risk that are protected and unprotected by structural flood defence systems, respectively. The other parameter corresponds to the proportion of EAD that is likely to be saved by each RRA shown in Figure 2.9. Table 2.3 shows the values suggested by Priest *et al.* (2011) to be used for each of them. These values were obtained from several social survey research projects conducted in England and Wales.

Table 2.3 Damage avoided for several RRAs.

RRA	Avoided damages due to early warnings [%]
Flood defence operations (FDO)	28
Watercourse capacity maintenance (WCM)	10
Community based operations (CBO)	1
Contingent resilience measures (CRM)	2
Contents moved and evacuated (CME)	5
Business continuity planning (BCP)	5

The methods mentioned above do not consider technical or social factors associated with the FEWS that can influence its economic benefits. The Environment Agency National Flood Warning Centre (2003) proposed a method that considers these factors in the economic effectiveness of flood warnings for the residential sector. This model was based on the work done by Parker (1991) and has been applied in England and Wales mainly to estimate the economic benefits associated with CME. This method gives a good insight into factors influencing the FEWS performance, which considers the following equation.

$$FDA = (EAD\ x\ DR\ x\ C) + (R\ x\ RA\ x\ PR\ x\ RE) \quad \text{Eq. 2.2}$$

Where FDA has already introduced above (Eq. 2.1), DR is the proportion of EAD that is likely to be saved by a pre-flooding action, C is the coverage of the flood warning service represented by the proportion of houses in the floodplain monitored by the system, R is the service effectiveness understood as the proportion of flooded serviced properties that received a timely, accurate and reliable flood warning, RA

is the availability of the flood warnings defined by the proportion of flooded services properties that received such a warning, PR is the ability of the floodplain dwellers to respond to a flood warning defined as the portion of residents able to understand and respond to such a warning, and RE is the effective action (proportion of serviced properties either willing to take effective action or which have actually taken effective action following a flood warning to reduce flood damages). In this research, PR and RE were encapsulated in a factor named the response human factor (RHF).

All the methods described above are often considered in a cost-benefit analysis of a FEWS where the costs of setting up, operating, and maintaining the system can be included (Pappenberger *et al.*, 2015). These approaches fail in that the cost of the warning response C_w (per-event costs) cannot be included in the cost-benefit analysis. It is a cost incurred every time a warning is issued and considers the cost of issuing a warning and any cost incurred in the dissemination process and by the RRA. It is, therefore, an important factor controlling the economic effectiveness of a FEWS. It depends on the forecasts because it is the only information available when the warning is issued. The cost of the warning response, C_w, is, therefore, affected by the accuracy of the forecasts, and its inclusion in the *economic effectiveness* of a FEWS is considered a difficult task. Verkade and Werner (2011) proposed a quasi-analytical framework to include C_w into the *economic effectiveness* of a FEWS by combining a hydro-economic EAD model with the theory of relative economic value. However, their framework has two main assumptions: i) The cost of the warning response is assumed to be independent of the magnitude of the forecast flooding, and it is not estimated as a function of this variate, and ii) cost and damage (net damage) associated with PU are only estimated as a result of "wrong" flood warning decisions (misses and false alarms). Their framework, therefore, negates the fact that the net damage associated with the PU can also be present in a hit event (a flood proceeded by a warning in the floodplain) ; this net damage is present in hits due to the difference between the warned and flooded houses

In Chapter 6, the economic effectiveness of a FEWS is quantified for a simulated fluvial case which considers the net damage associated with PU *in all possible situations.* This analysis, to the best of the author's knowledge, has not been done before. In the case study (a FEWS operating for a flood-prone polder system),

Chapter 7, the proactive action benefits are jointly analysed together with their costs.

2.8 Operational FEWS

This section of the Chapter compiles relevant information about real operational FEWSs to describe how these real-world FEWSs are actually operated and evaluated. It is important to understand the relevance and motivation of the research. The actual operation of FEWS can be summarized as follows.

- ✓ **Lead time:** A hydro-meteorological modelling chain - QPE/QPF plus a hydrological model - for long-term flood forecasts (1-5 days) have not been used very extensively due primarily to that: i) the technology for QPE/ QPF procedures are not available and applied easily in many countries and regions; ii) the lack of techniques that include hydrological models in the calibration/validation of the QPE/ QPF products; and iii) its inadequate spatial and temporal resolution for hydrological modelling purposes (Liu, 2012). Thus, long-term flood forecasts have been generated by local FFS through meteorological forecasts, mainly QPF, (global- or national-scale products) often provided by meteorological agencies in the form of NWP (Rabuffetti and Barbero, 2005; Dugar *et al.*, 2017; Flack *et al.*, 2019). However, these forecasts have been used only qualitatively, i.e., to track the threat of flooding and sometimes to warn only flood authorities (not specific risk zones) to prepare the warning response. Thus, in practice, real-time FFSs have been used in FEWS and, therefore, flood warning decisions have been based on forecasts based on real-time hydro-meteorological information (Arheimer *et al.*, 2011; Adams, 2016; Javelle *et al.*, 2016; Liu *et al.*, 2018; BOM, 2019). That has caused FEWSs to consider short lead times (lower than the hydrological lead time), especially for small to medium size catchments. For example, in England, the EA issues flood warnings with a *forecasting lead time* between 30 min and 6 hours (Arnal *et al.*, 2020).

- ✓ **Forecast variates**: The timing and the peak of the floods, along with the inundation extent, at key locations, are the most relevant variables to be forecasted (Ibbitt and Woods, 2003; WMO, 2011; Parker, 2017; Jain et al., 2018). From these variates, the peak and timing of floods are often forecasted because FEWSs often predict floods based on discharge or water levels; the inundation extent is rarely forecasted (Fernández-Nóvoa et al., 2020). Some FEWSs also predict the flood peaks based on rainfall, where the rainfall threshold (RT) method (Golian et al., 2015) is a common choice.

- ✓ **Forecast type and warning decisions:** Flood warning decisions have been based on deterministic forecasts, though probabilistic forecast is starting to be considered in operational FEWSs due to the significant progress in the last two decades in the assessment and estimation of uncertainty in flood forecasting through ensembles (see the case for England in Arnal *et al.*, (2020)). The warning-decision-making process is complex, and forecasts represent only one factor influencing this decision. Factors such as expert

judgement, type of event, risk attitude, and cultural environment in which decisions are made also influence the warning decision (Ramos *et al.*, 2010; Verkade and Werner, 2011; Arnal *et al.*, 2020).

- ✓ **Common methods to evaluate performance:** Operational FEWSs have been often evaluated in terms of the *reliability* of flood warnings and their *economic effectiveness*. The former attribute is often evaluated based on whether a warning preceded a flood event's occurrence or non-occurrence in the floodplain *or* in a specific flood risk zone, where a pre-defined threshold defines the occurrence of the event. The metrics such as POD, FAR, F, or CSI are often used to have a "snap-shot" of the flood warnings' reliability in the floodplain or a specific flood risk zone (section 2.7.1). Operational FEWSs often compute these metrics based on a flood warning validation database. This is a common practice in, for example, England (Arnal et al., 2020). These metrics do not consider the uncertainty in the estimation of the flood magnitude and, therefore, the difference between the warned and flooded properties. Quantifying the *economic effectiveness* of an operational FEWS is, on the other hand, considered a challenging task due to the complex nature of the warning response processes (Pappenberger *et al.*, 2015; Girons Lopez *et al.*, 2017). Thus, operational FEWSs rarely have information about the economic benefits they may produce (Economou *et al.*, 2016). Most studies in this field have estimated this attribute by doing the ex-post evaluation of the system's benefits based on interviews or surveys. Several methods to estimate the *economic effectiveness* of a FEWS have emerged from these studies, which often neglect the net damage associated with the accuracy of a forecast (e.g., costs of false alarms and damage resulting from misses).

Based on the operational characteristics of the real FEWS mentioned above, it would not be easy, but not impossible, to obtain relevant information associated with the influence of the factors described through this Chapter and Chapter 1 on the FEWS performance. This research considers flexible Monte Carlo frameworks for these purposes that consider the end-to-end modelling of a FEWS. In this sense, the next section of this chapter describes research works that have done the end-to-end simulation of these systems. This research takes some concepts of them to build a generic framework (explained in Chapter 3), which is the basis to do the investigations of this research.

2.9 End-to-end modelling of FEWS

Few research projects have considered an end-to-end-modelling framework to analyse the performance of a FEWS. Section 1.2 provides a discussion of about several important research works in this field, whereas Table 2.4 a summary of their aims and their main assumptions. Beyond each particular work's specifics, all of these works have idealized the FEWS through the forecast-decision-

response/impact chain illustrated in Figure 2.10. By doing that, they have simplified several processes considerably involved in the integrated system obtaining, thus, more or less versatile frameworks that meet their goals. Even though these works provide good insights into the representation and evaluation of the performance of a FEWS, none of them i) have explored the impact of the flood magnitude's uncertainty on the FEWS performance, ii) have explored the impact of the bias in the variance on the reliability of flood warnings, iv) have explored the impact of several factors on the economic effectiveness of a FEWS, and v) have used the forecast-decision-response/impact chain to represent a FEWS operating for a flood-prone polder system.

Table 2.4 Important research works that have addressed the FEWS performance through an end-to-end modelling framework.

Research work	Aim	Representation of the FEWS and main assumptions
Krzysztofowicz and Davis (1983)	To evaluate the economic effectiveness of FEWS	The FEWS chain is reduced to a flood forecast-response process; warning decisions are not considered
Krzysztofowicz (1993)	To evaluate the reliability and economic effectiveness of a FEWS	The FEWS is represented through a monitor-forecast-decision chain; the response is not simulated.
Verkade and Werner (2011)	To evaluate the economic effectiveness of a (probability and deterministic) FEWS based on the relative economic value (REV) theory.	FEWS is represented through a forecast-decision-response-and-impact chain. REV is used to scale the imperfect-forecast-based-economic flood risk between the economic flood risk of the perfect and no warning scenario; the imperfect forecast is not explicitly simulated.
Girons Lopez *et al. (*2017)	To explore the impact of social preparedness on the economic effectiveness of a FEWS	FEWS is represented through a forecast-decision-response-and-impact chain; the impact of the lead time on the FEWS performance is not explored.
Bischiniotis *et al.* (2019)	To explore the decision-makers' dilemma between acting upon limited-quality forecast information and taking less effective actions	FEWS is represented through a forecast-decision-response-and-impact chain; results are not based on a considerable range of potential flood events *and* warning decision situations that the area monitored by the FEWS may be subject to.

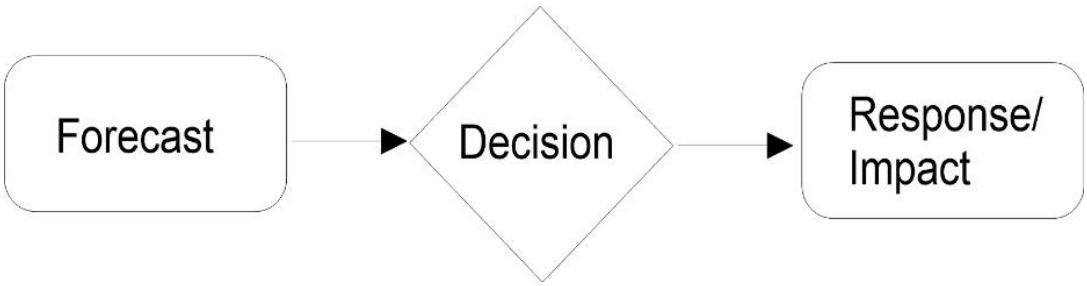


Figure 2.10: Forecast-decision-response/impact chain often used to analyse FEWS
Adapted from Verkade (2019).

This research uses the forecast-decision-response/impact chain and Monte Carlo simulation to fill the research gaps mentioned above. Monte Carlo simulation has for many decades been used to address uncertainties in hydrologic series and the impact these have on decision making. In the seminal work of Hosking and Wallis (1997), it was demonstrated how Monte Carlo sampling can be used to explore the sensitivity of the estimate of T-year flood to various factors and assumptions. However, few studies have utilised Monte Carlo simulation in flood forecasting. Examples include Van Steenbergen and Willems (2014), who investigated whether ensemble predicting systems (EPS) cover the uncertainty produced by Numerical Weather Prediction model rainfall forecasts, and Golian *et al.* (2010), who developed a probabilistic rainfall threshold curve for flood forecasting. Chapter 3 describes how this research addresses these issues and the methodology used.

2.10 Main Findings

Based on the research gaps identified in this literature review, flexible Monte Carlo frameworks have been designed to do the end-to-end modelling of a FEWS operating for a fluvial-flood-prone area (Chapters 5 and 6) and a flood-prone polder system (case study, Chapter 7). These frameworks allow simulating some of the factors influencing the performance of a FEWS identified through this Chapter.

Table 2.5 summarises all of them according to the component they correspond to in the FEWS. Factors as the failure of instruments/telemetry system (FITS), the failure of communication networks (FCN), the flood warning service effectiveness (R), and the coverage of the flood warning service (C) are not considered in this thesis. Likewise, human errors in the forecast and warning process (HFWP) are also ignored. These omissions are supported by the fact that perfect operability and communication and coverage of 100% of the flood warning service is assumed in the frameworks designed to simulate the FEWSs. The human component in the

flood warning decision-making process (HFWD) is also ignored since it is assumed that warning decisions are made based on pre-stated decision rules.

Table 2.5 Main factors influencing the performance of a FEWS

Symbol or acronym	Description	Factor associated with:
HFWP	Human errors in the forecast and warning process	Flood forecasting and warning system
HFWD	The human component in the flood warning decision-making process	
LT	Lead time	
FBM	Forecast bias in the mean	
FBV	Forecast bias in the variance	
FITS	Failure of instruments/telemetry system	
FCN	Failure of communication networks	
R	Flood warning service effectiveness	
C	Coverage of the flood warning service	
C_w	Cost of the warning response	
HRF	Human response factor	Response system and the at-risk area
CRRA	Cost of the RRA	
SoP	Standard of protection	

Factors influencing the performance of the FEWS considered in this research were forecast bias in the mean (FBM) and variance (FBV) of the predictand, the cost of the warning response (C_w), the cost of the RRA (CRRA), the lead time (LT), and the standard of protection (SoP). On the other hand, only one human factor was considered; the response human factor (RHF), which, as was mentioned in section 2.7.2, includes the ability of the floodplain dwellers to respond to a flood warning and the effectiveness of the RRA. These factors have been parametrised in flexible Monte Carlo frameworks designed in this research. Thanks to the versatility of these frameworks, this research obtains relevant information associated with the influence of these factors on a FEWS performance based on several scenarios associated with the setting of the systems. Part of this information represents a contribution of this research, and it would not be easy to be derived from information provided by real-world FEWSs due to its operational characteristics

detailed in section 2.8. This information was generated by answering the following main research questions:

- ✓ What levels of correlation between observed flood and their forecasts are needed to obtain target levels of reliability of flood warnings, and how this varies with the lead time? (Chapter 5)
- ✓ How sensitive is flood warning reliability to biases in the mean and variance of deterministic and probabilistic forecasts? (Chapter 5)
- ✓ How is the economic effectiveness of FEWSs affected by various factors?

Other questions that do not involve these factors but can be answered with the adopted methodology are:

- ✓ How should the impact of the flood magnitude's uncertainty on FEWS performance be explored?(Chapter 5)
- ✓ How does the uncertainty in forecasts and its quantification through PU impact the performance of a FEWS? (Chapter 5, 6, and 7)
- ✓ Can the management of floods in polder areas be improved through the use of flood forecast and warnings, and how the response component be designed? (Chapter 7)

The answer to these questions and the MC frameworks considered to fill out the research gaps mentioned above are described through the thesis and highlighted in the thesis conclusions (Chapter 8).

Chapter 3. Methodology

A generic framework has been designed to address the research needs identified in Chapter 1 and highlighted in Chapter 2. Here, the aim is to provide an overview of the framework, to introduce the components that are developed in the following chapters, and to show how they relate to each other. This can be viewed as a thesis roadmap, thus helping the reader to navigate the thesis. In addition, after the introduction of this generic framework, this Chapter closes by describing the type of sensitivity analysis (SA) adopted in this research to identify the sensitivity of important factors controlling the performance of the FEWSs considered in this thesis, which is one of the primary research objectives of this research.

3.1 The generic framework

The generic framework has been designed based on the forecast-decision-response/impact chain used to characterize the end-to-end modelling of a flood early warning system (FEWS) (Figure 2.10). When it was designed, it was thought that it should represent this chain through sampled Monte Carlo realizations of observed values and their forecasts, the resulting decision(s) taken, and the impact of the observed realizations on the floodplain, conditional on the decision(s) taken. Based on this approach, the resulting generic framework is made up of three components, as illustrated in Figure 3.1. It is explained as follows.

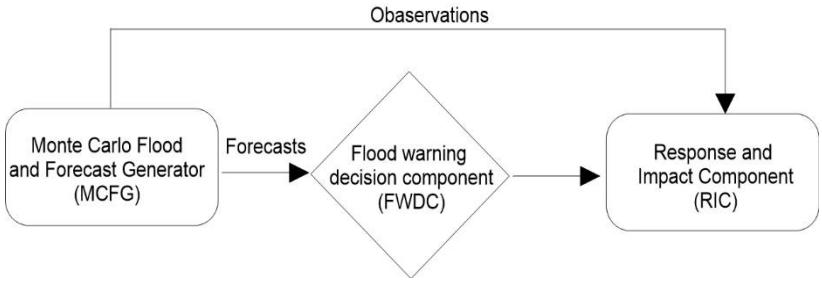


Figure 3.1: Generic framework of this research

- ✓ **Monte Carlo flood and forecast generator (MCFG)**: It is used to generate the full range of flood forecasts in terms of deterministic and probabilistic forecasts and their observed realizations for an at-risk community.

- ✓ **Flood warning decision component (FWDC):** It represents the warning decision, and it is simulated by decision rules which represent rational decision-making. The FWDC is driven by the forecasting information generated by the MCFG.

- ✓ **Response and impact component (RIC):** This component is used to simulate the proactive action and the resulting impact. The proactive action

is driven by forecast information and warning decisions, whereas the impact is estimated based on the observed realizations generated from the MCFG and conditional on the warning decisions.

3.2 Component description

The generic framework depicted in Figure 3.1 has been used to explore the performance of a FEWS operating for a generic fluvial flood case (Chapters 5 and 6) and a real flood-prone polder system (Chapters 7). Event-based modelling and continuous simulation are used in the former and latter cases, respectively; the MCFG is used to sample the driving forecasts and their observed realizations (Chapter 4). The parameters of the components vary according to the research objectives addressed and the assumptions considered to explore them; they are supported by the analysis of observed data.

3.2.1 The Monte Carlo Flood and Forecast Generator (MCFG)

The MCFG assumes a bivariate parametric probabilistic model for the predictand (the observed value) and predictor (the forecast) and assumes perfect knowledge about its parameter values, which are varied to explore the FEWS performance sensitivity to different factors. Under this assumption, the MCFG generates bivariate values of these two variates based on the method known as the conditional approach (Lewis and Orav, 2018). Thus, it first obtains the conditional probability density of the predictand given a value of the predictor; and, then, the (observed) realization of the predictor is obtained as a value sampled from this conditional probability distribution. The predictor's value represents the deterministic forecast, and the probabilistic forecast is derived from the conditional probability distribution. Even though the aim of the MCFG for the simulated generic fluvial case and flood-prone-polder system case is the same, i.e., to generate potential long-term scenarios of flooding in an at-risk community, the predictand and predictor considered, and the architecture of the MCFG, is different in each of them.

Generic Fluvial Case

In the simulated fluvial case (Chapters 5 and 6), a potential flood in the at-risk community is defined by observed peak water levels (y). Thus, here, y is assumed to be the predictand, and its forecast ($\hat{y}$), the predictor. The bivariate parametric probabilistic model is, therefore, defined by the parameters that describe the

marginal distributions of y and $\hat{y}$, and the correlation coefficient ($\rho_{y\hat{y}}$) that defines the dependence structure of the pairs (y,$\hat{y}$). The MCFG in the generic fluvial case, therefore, generates a deterministic-single-value forecast, represented by the values of $\hat{y}$, and its associated probabilistic forecasts derived from the conditional distribution of y given $\hat{y}$, i.e., $f(y|\hat{y})$.

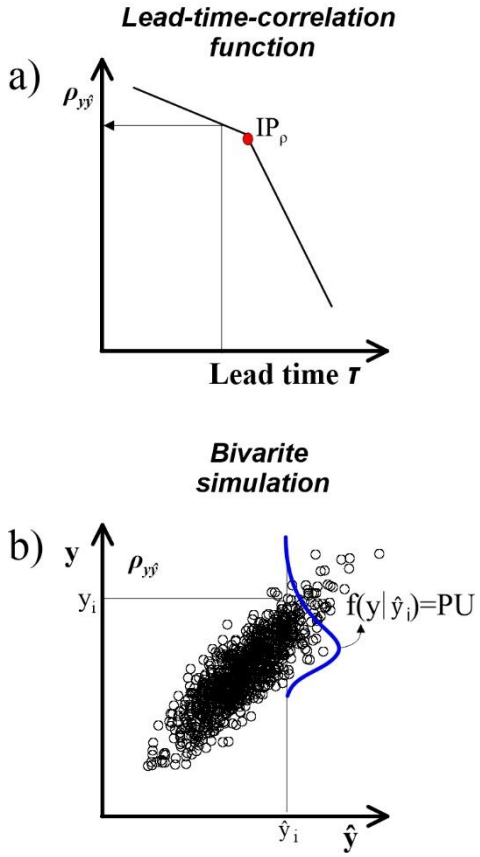


Figure 3.2: Schematic of the MCFG for the simulated fluvial case
This figure illustrates the bivariate modelling of pairs (y,$\hat{y}$) from a pre-assumed parametric bivariate model (b), where the value of $\rho_{y\hat{y}}$ of this bivariate model is associated with a given value of lead time τ and derived from the subjective lead-time-correlation function (a).

The correlation coefficient $\rho_{y\hat{y}}$ in the bivariate parametric model, in essence, establishes the scatter of the bivariate points and is one of the parameters of the bivariate model that define the *predictive uncertainty* (PU). In this sense, the MCFG, in the simulated fluvial case, uses a subjective lead-time-correlation function to associate $\rho_{y\hat{y}}$ with a given forecast horizon or *forecasting lead time* (τ). This function is analogous to that found by Schröter *et al.* (2008) and aims to represent the trade-off between τ and PU. This function, depicted graphically in Figure 3.2a, describes the common behaviour of forecasting models forced with precipitation. That is, it represents the fact that, for τ values lower than catchment

lag time (L), with a well-calibrated hydrological model, and forecast updating in real-time, the performance of a forecasting model is relatively high as forecasts are based on observed precipitation by using, for example, gauge-based quantitative precipitation estimation (QPE). Past this value L, the forecasting model has to be forced with quantitative precipitation forecasts (QPF)(Figure 2.2), and its performance is hypothesized to drop monotonically (Schröter *et al.*2008). The slope of this function before L defines the quality of forecasting models based on QPE, such as models based on gauge-based QPE or gauge-radar-based QPE, and the slope after L defines the quality of the forecasts based on QPF. The architecture of the MCFG in the simulated fluvial case is, therefore, made up of the bivariate parametric model of the pairs (y,$\hat{y}$) *and* the subjective lead-time-correlation function. Figure 3.2b illustrates the bivariate modelling of pairs (y,$\hat{y}$) from a pre-assumed parameteric bivariate model, where the value of $\rho_{y\hat{y}}$ of this bivariate model is associated with a given value of τ and derived from the subjective lead-time-correlation function. In this sense, the MCFG in the fluvial case generates (deterministic and probabilistic) single-value forecasts associated with a τ value, which are provided to the FWDC, and their single-value- observed realizations, which are delivered to the RIC.

Polder Case

In the simulated flood-prone-polder system case (Chapter 7), a flood in the polder area is predicted based on forecasting rainfall over a 24-hour forecast horizon (*forecast lead time*). A Rainfall Forecast Generator (RFG) is constructed that first uses the space-time rainfall generator RAINSIM v3 to generate 24-hour forecasts and observed values at hourly resolution with a prescribed correlation using the model in spatial mode; here, spatial correlation is used as a surrogate for the correlation between forecast and observed values. Then, to obtain a probabilistic 24-h rainfall forecast, a bivariate model is constructed based on the daily observed rainfalls R_{daily} and their forecasts $\hat{R}_{daily}$from RAINSIM. Thus, here, the daily rainfall (R_{daily}) is assumed to be the predictand, and its forecast ($\hat{R}_{daily}$), the predictor. The bivariate parametric probabilistic model is, thus, defined by the parameters that describe the marginal distributions of R_{daily} and $\hat{R}_{daily}$, and the correlation coefficient ($\rho_{R_{daily}\hat{R}_{daily}}$) that defines the dependence structure of the pairs (R_{daily},$\hat{R}_{daily}$).Therefore, the MCFG, in this case, generates (deterministic

and probabilistic) 24-h rainfall forecasts, which are provided to the FWDC, and their 24-h hourly observed realizations, which are delivered to the RIC to perform continuous hourly simulations of the polder operation over a 24-hour storm period.

A detailed description of the potential bivariate parametric probabilistic models one can use to represent the pairs $(y,\hat{y})$ is provided in Chapter 4, together with supporting data analysis for some UK rivers. The data analysis supporting the choice of a bivariate distribution for $(R_{daily},\hat{R}_{daily})$ is described in Chapter 7.

3.2.2 The flood warning decision component (FWDC)

The FWDC is used in the framework to simulate the warning decision of a FEWS. The warning decision, driven by the forecast, is represented by decision rules set based on the warning criteria adopted in the analysis of the FEWS.

Generic Fluvial Case

In the simulated fluvial case, this research considers FEWSs based on a flooding threshold (y_T) with *and* without real-time flood maps. Figure 3.3 illustrates the deterministic and probabilistic rules used in the simulated generic fluvial case. Chapter 5 uses these rules to explore the flood warning reliability of these two types of FEWs. In the deterministic forecast scenario, the warning decisions in these two systems assume that a warning is automatically issued when the single-value forecast is greater than the warning threshold y_T which is determined by the T-year standard of protection (SoP). In the probabilistic forecast scenario, warning decisions are simulated through a probabilistic threshold (PT) approach. In the FEWS *without* real-time flood maps, probabilistic warning decisions assume that a warning is automatically issued when the probability of exceedance (PE) of y_T conditional on the forecast exceeds a pre-defined probabilistic threshold PT (Chapter 5). In the FEWSs *with* real-time flood inundation maps, probabilistic warning decisions assume that a warning is automatically issued when a warning level ($\hat{y}_w$) defined from $f(y|\hat{y})$ is greater than y_T, where $\hat{y}_w$ is also defined by a PT. In both types of FEWS (*without/with* real-time flood maps), PT is a value to be optimized by assuming that the Warner acts to increase the *reliability* of flood warnings evaluated based on the flooding threshold-based criterion (FTC) for the FEWS *without* real-time flood maps, or floodplain property-based criterion (FPC), for the FEWS *with* real-time flood maps (see section 2.7.1 to recall the difference

between FTC and FPC). In Chapter 6, the *economic effectiveness* of flood warnings based on real-time flood maps is explored, and those decision rules used for this type of FEWS in Chapter 5 are also considered there. However, in this case, the PT associated with the probabilistic rule is optimized by assuming that the Warner acts to increase the *economic effectiveness* of flood warnings.

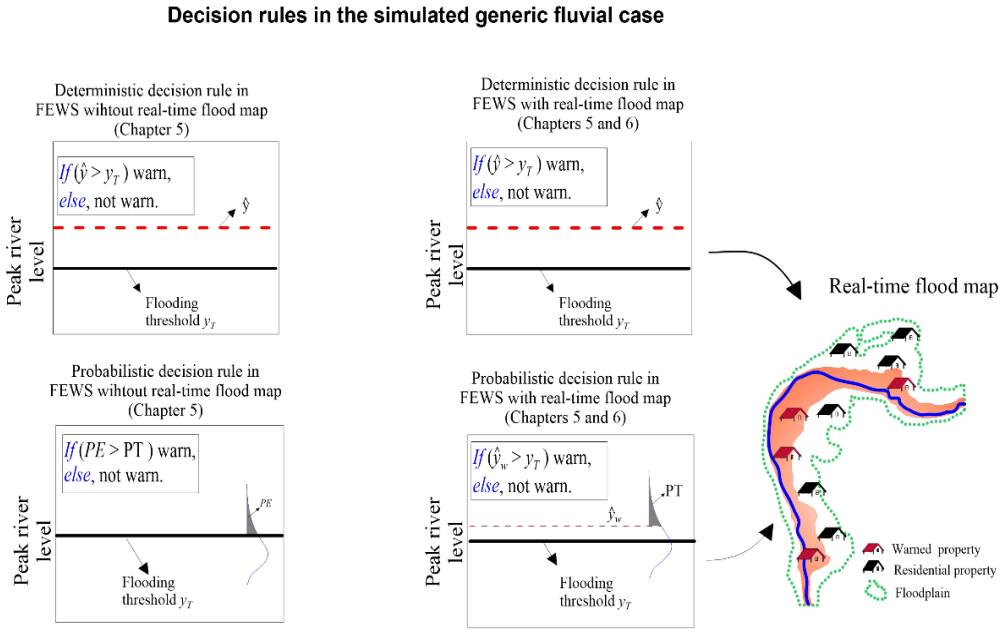


Figure 3.3: Description of the FWDC for the simulated fluvial case

To analyse the flood warning reliability of a FEWS through the FPC, one needs to distinguish between houses that have been warned and those which have been flooded, and POD and FAR are estimated from these quantities. In this context, the framework does not simulate the generation of the real-time forecast and observed flood maps explicitly but simulates their target information, i.e., the number of warned houses (n_{houses}^{warned}) which is determined by $\hat{y}_w$, and the number of flooded houses which is determined by y where $y > y_T$. The numbers of warned n_{houses}^{warned} and flooded $n_{houses}^{flooded}$ are obtained by interpolating the values of $\hat{y}_w$ and y, respectively, in an assumed impact curve, which gives an estimate of the percentage of floodplain properties affected by different floods (Figure 3.4). The impact curve assumes that when the SoP is overtopped, and water spills into the flood plain, houses start to be flooded above $y_{bankfull}$.

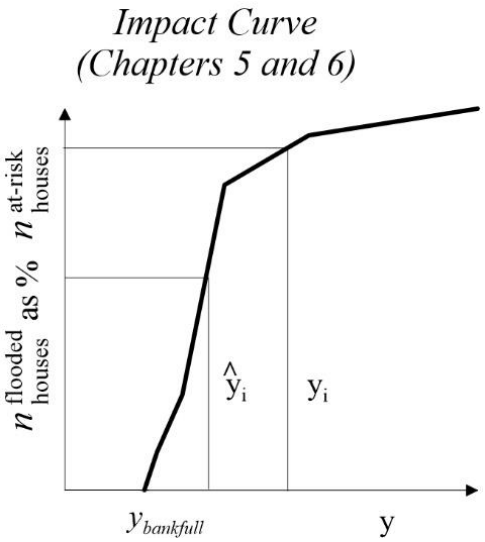


Figure 3.4: Description of the FWDC for the case study

This impact curve assumes that the flood impact starts when the bankfull level is overtopped. The figure illustrates the computation of n_{houses}^{warned} and $n_{houses}^{flooded}$ for a pair $(y, \hat{y})$.

Polder Case

Figure 3.5 illustrates the deterministic and probabilistic rules used in the polder case study. A rainfall threshold (RT) curve is used as a tool for issuing flood warnings. The RT curve defines critical volumes of daily rainfall on the polder area that bring the inner rivers to the critical condition. This curve is made up of different critical values associated with several initial conditions of the water level of the inner rivers (h_0) at the time the forecast is issued. Thus, in the deterministic forecast scenario, it is assumed that warnings are automatically issued when the deterministic 24h-forecasts, i.e., $\hat{R}_{daily}$, cross the RT curve. In the probabilistic forecast scenario, probabilistic warning decisions assume that a warning is automatically issued when the PE of the RT curve exceeds a pre-defined probabilistic threshold PT. PE is obtained from $f\left(R_{daily}|\hat{R}_{daily}\right)$, and PT is a value to be determined when analysing the performance of the FEWS in terms of the costs of the response.

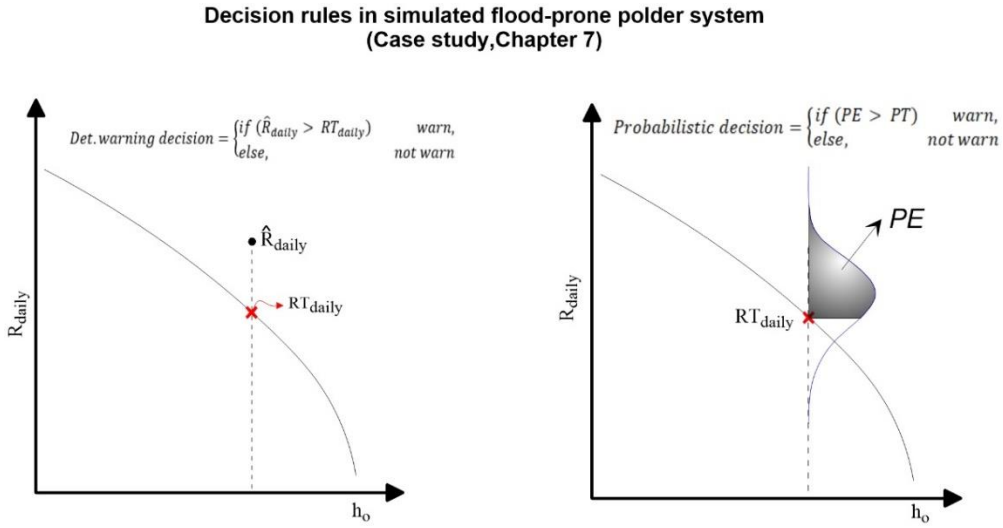


Figure 3.5: Description of the FWDC for the case study

3.2.3 The response and impact component (RIC)

After the warning decision is made, the response and the occurrence or non-occurrence of the flooding, along with its impact, have to be simulated. These processes are simulated in the RIC.

Generic fluvial Case

In Chapter 6, the *economic effectiveness* of flood warnings based on real-time flood maps is analysed. In this case, the RIC is used to estimate the flood impact and the response to flood warnings. The observed flood impact is estimated in terms of $n_{houses}^{flooded}$ and flood depth. The former variate is estimated based on the impact curve-based approach adopted in Chapter 5 to estimate the flood warning reliability in terms of FPC, whereas the flood depth in the houses is computed as the difference between the magnitude of the flooding (defined by the values of y greater than y_T) and the bankfull level ($y_{bankfull}$). The economic flood damage in each floodplain property is estimated by using a damage function, which represents the damage to the residential content as a function of the *mitigation time* (which is assumed in this research to be equal to the *forecasting lead time*) and flood depth. A schematic of a damage function associated with a specific mitigation time is shown in Figure 3.6. A family of these curves prepared by Carsell et al. (2004) for a sample of US houses and range of mitigation times is used in Chapter 6. The RIC is also used to *emulate* the real-time forecast of flood damage. This information is obtained based on the same approach adopted for the observed number of houses flooded. In this case, the

forecast values of y are used to estimate n_{houses}^{warned}, the forecast flood depth, and the associated flood damage.

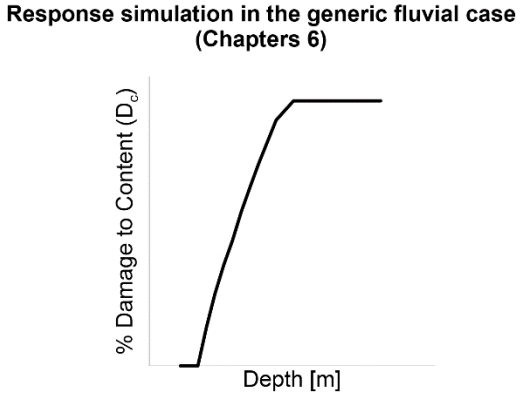


Figure 3.6: Schematic of a damage function used in the RIC for a specific mitigation time in the generic fluvial case

The RIC simulates the response of householders to flood warnings through the functions found by Carsell *et al.* (2004).

Polder Case

The RIC in the polder case (Figure 3.7) considers a polder system where pumping systems operate before or during storm events to remove water from the inner rivers to the outer rivers in a selected polder area of Nanjing, to enable water to drain from the polder areas into the inner rivers. In that case, the response represents a pumping scheme operating in a polder system, and the impact is estimated as the resulting area of waterlogging (W) in the polder after the pumping action has been done. These two processes are simulated in a lumped water balance model that uses an assumed impact curve to estimate the inundated area as a function of W. A lumped water balance model is used to represent the response of the polder to rainfall and to represent the effects of the proactive and reactive pumping actions (responses). The forecasts and the warning decisions drive the responses, and, therefore, they start before the storm arrives in the polder area. The reactive pumping actions start when the storms arrive in the polder area and are driven by the inflow entering the polder's inner rivers. The tradeoff between the waterlogged area and the pumping cost is explored as a function of the proactive and reactive actions and both the deterministic and probabilistic forecast scenarios.

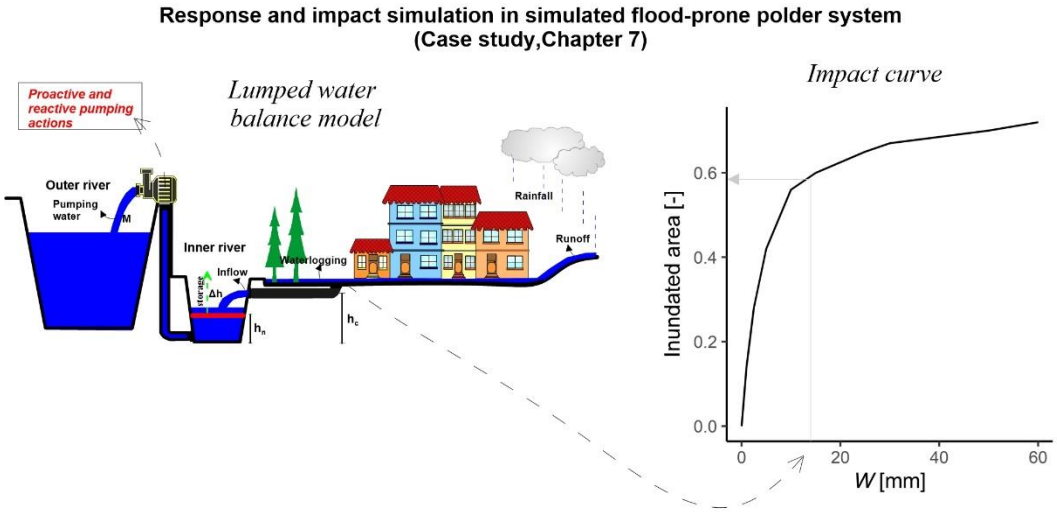


Figure 3.7: Description of RIC in the polder system case study

3.2.4 Benchmark scenarios

When evaluating a FEWS in economic terms, it is customary to simulate the no warning scenario and the perfect forecast scenario. The aim is to compare the economic performance of the FEWS, assuming the warning service does not exist in the area monitored by the FEWS, or assuming perfect knowledge of the future values. Thus, these two scenarios are used as benchmark cases to analyse the 'location' of the economic performance of the imperfect FEWS between them. By definition, the perfect forecast scenario results are never outperformed by the results of the imperfect FEWS, but the results of the no warning scenario may be better than the imperfect-forecast-based results due to the economic consequences of high PU. Benchmark scenarios are used in Chapters 6 and 7. Since there are no forecasts in the no warning scenario, the FWDC is removed in this case

Generic fluvial Case

For the perfect forecast scenario and no warning scenario in the simulated generic fluvial case (Chapter 6), potential floods in the at-risk community are simulated from the marginal distribution of y of the bivariate parametric model considered in the MCFG. The flood depth is estimated by the approach described above for the imperfect forecast FEWS.

For the no warning scenario, $n_{houses}^{flooded}$ is simulated through the assumed impact curve, and the flood damage in each floodplain property is estimated by using the functions proposed by Carsell *et al.* (2004) and assuming a mitigation time of zero. These functions are also used to estimate $n_{houses}^{flooded}$ and the economic damage in the

perfect forecast scenario, which assumes that each warned house is subsequently flooded, i.e., $n_{houses}^{flooded}$=n_{houses}^{warned}.

Polder Case

In the simulated flood-prone-polder system case, the no warning scenario is simulated by considering only reactive pumping actions in the lumped water balance model. The perfect forecast scenario is simulated by considering only proactive pumping actions in the lumped water balance model and assuming perfect knowledge of the storm's profile and volume.

3.3 Performance measure estimation

As one can see, the generic end-to-end modelling framework allows forecast outputs to be linked with warning decisions and the associated responses and flood impacts. Thus, based on a representative sample of (deterministic or probabilistic) forecasts and their observed realizations, the average performance of a FEWS can be synthetized based on metrics computed from the outputs of the end-to-end simulation of the system associated with the sample.

For the simulated fluvial case, the metrics used to analyse the reliability of a FEWS are the probability of detection (POD) and false alarm ratio (FAR) for each lead time (Table 2.2), while the economic metric used is the expected damage (ED) derived by taking the convolution of the frequency distribution of flood depth with the house damage function for each lead time. The *reliability* of flood warnings is explored through FAR and POD estimated based on the FTC and FPC criteria in Chapter 5, whereas ED values are used in Chapter 6 to define the *economic effectiveness* of flood warnings relative to the no warning scenario. In the simulated flood-prone-polder system case, metrics such as the average of the maximum inundated area ($\overline{MIA}$) and the waterlogging duration ($\bar{d}_w$) are considered for evaluating the performance of the FEWS. Since these performance measures are related to the pumping costs, the average pumping costs ($\overline{PC}$) are also included in the analysis.

3.4 Sensitivity analysis

One of the primary research objectives of this thesis is to explore how the performance of a FEWS behaves under different factors controlling the performance of the system. Since these factors have been parametrised in the Monte Carlo frameworks that represent the FEWSs considered in this research, this overall research aim of the thesis was addressed by conducting a (sensitivity analysis) SA of the parameters representing these factors. The methods commonly adopted for doing a SA can be classified as two: local and global (Pianosi *et al.*, 2016; Devak and Dhanya, 2017). Local SA is a one-at-a-time (OAT) technique that, in essence, analyses the effect of one parameter change at a time on the model being evaluated, keeping the other parameters fixed. Global SA, on the other hand, considers variations within the entire space of variability of the parameters. Since the main aim in the SA was to do a first-order analysis that would reveal the main sensitivities of the factors controlling the FEWS performance, a local SA analysis was adopted. Thus, this type of SA was conducted in the simulated generic fluvial case (Chapter 5 and 6) and flood-prone polder system case (Chapter 7).

Chapter 4. Monte Carlo Flood Peak and Forecast Generator (MCFG)

4.1 Introduction

This document gives a detailed description of the first component of the generic framework of the research; the Monte Carlo flood and forecast generator (MCFG) (see Figure 4.1). The MCFG is associated with a given forecast horizon or lead time (τ) and is based on a bivariate probabilistic model of the peak flows (y) and their forecasts ($\hat{y}$) whose marginals are represented by appropriate distributions. Here, the generic fluvial case is taken as a reference. The deterministic and probabilistic forecasts, and a measure of predictive uncertainty (PU), are derived from that model.

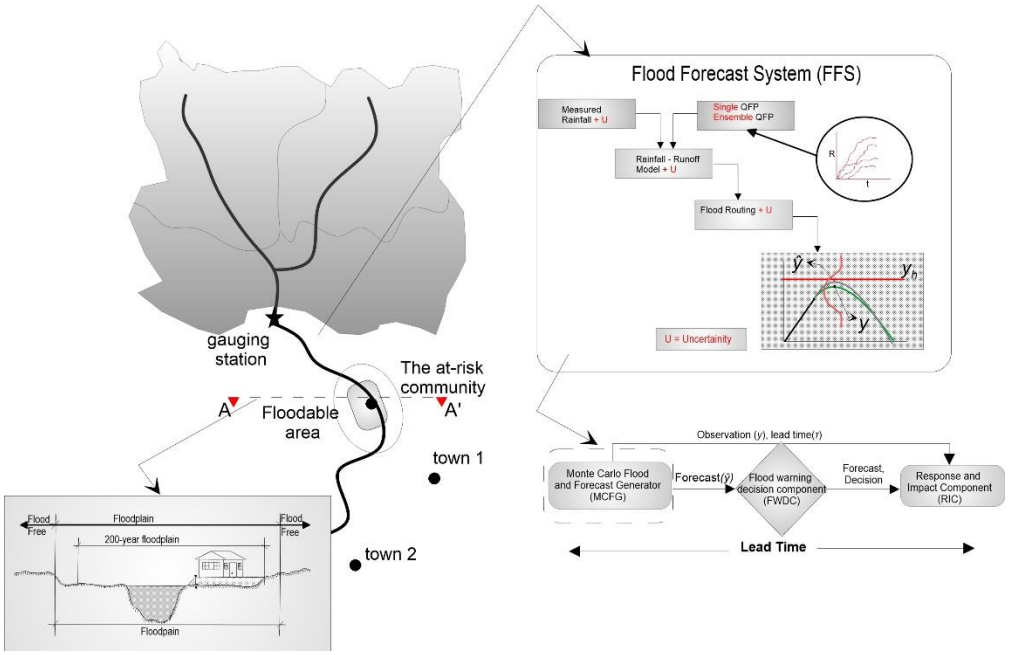


Figure 4.1 Illustration of a type of forecasting model to be represented in the generic fluvial case. The figure shows a potential configuration of a forecasting system operating for an at-risk community when the lead time (τ) is bigger than the response time of the catchment. In this case, the hydrological model is forced with quantitative precipitation forecasts (QPFs).

Thus, this Chapter is organized as follow: Section 4.2 gives a description of the probability distributions considered as candidates to represent y and $\hat{y}$; section 4.3 describe the theory used when doing the bivariate modelling of these two variates and shows the algorithms built to simulate pairs (y,$\hat{y}$) and to build the PU assuming they can be represented by a five-parameter bivariate Exponential distribution (BED) and a seven-parameters bivariate Log-normal distribution (BLND). Section

4.5 presents data analysis which validates some assumptions made when choosing the shape of the marginal distributions of y and $\hat{y}$, and when choosing the bivariate model of these two variates. Section 4.6 explains the criteria used to associate the dependence structure of the pairs (y,$\hat{y}$), embedded in the bivariate probabilistic model, with τ and a given forecasting performance measure. Finally, section 4.7 shows the parametrization of the MCFG and the sensitivity analysis (SA) strategy adopted to define the parameters of the MCFG, which are considered in the SA of the performance of a flood early warning system (FEWS).

4.2 Univariate distributions

It is well-known that Exponential and generalized Pareto (GP) distributions can be used to simulate peaks through the peak over threshold-method (POT-method) (Claps and Laio, 2003; Bezak *et al.*, 2014). However, some works suggest that this variable can also be simulated through the Lognormal (LN)(Adamson and Zucchini, 1984), Gamma (Bačová-Mitková and Onderka, 2010), or the Generalized Logistic (GL) distributions (Bhunya et al., 2012). In this work, three of these distributions were considered as candidates for describing real data: the two-parameter exponential distribution, the three-parameter log-Normal distribution, and the three-parameter Gamma distribution. Since y and $\hat{y}$ will be analyzed based on the same probability distributions, to avoid too many different symbols, the following notation is used throughout this section:

- ✓ w: variable used to represent y or $\hat{y}$.
- ✓ w_o: variable used to represent the location parameter of y or $\hat{y}$, i.e. y_o or $\hat{y}_o$, respectively.
- ✓ z: for the two-parameter Exponential distribution and three-parameter Gamma distribution, it represents the transformed variables: $x = y - y_o$ or $\hat{x} = \hat{y} - \hat{y}_o$. For the case of the three-parameter Lognormal distribution, it represents the natural log of x or $\hat{x}$.

A summary of the equations of the cumulative density function (CDF), mean, and variance for the three probabilistic models considered in this work as candidates to represent y and $\hat{y}$ is shown in Table 4.1. A brief description of each of them, including the moment equations, is detailed as follows.

Table 4.1 Equations of the CDF, mean, and variance of the probabilistic models considered as candidates to represent y and $\hat{y}$.

The equations of the random variables Y and $\hat{Y}$ can be obtained by replacing w with y and $\hat{y}$, and, for the case of the Exponential and Gamma distributions, z with x and $\hat{x}$.

CDF	Mean	Variance
Two parameter-Exponential distribution $F(w) = \begin{cases} 1 - exp\left(-\frac{w - w_o}{\beta_z}\right), w \geq w_0, \\ 0, \quad w < w_0 \end{cases}$	$w_o + \beta_z$	β_z^2
Three-parameter Log-normal distribution $F(w) = \Phi\left(\frac{ln(w - w_o) - \mu_z}{\sigma_z}\right), w > w_0$	$w_o + exp(\mu_z + 0.5\sigma_z^2)$	$exp(\sigma_z^2 + 2\mu_z)[exp(\sigma_z^2) - 1]$
Three-parameter Gamma distribution $F(w) = \frac{1}{\Gamma(k_z)}\gamma\left(k_z, \frac{w - w_o}{\beta_z}\right), w > w_0$	$w_o + k_z\beta_z$	$k_z\beta_z^2$

Notation:

w: Variable to represent y or $\hat{y}$.

w_o: Location parameter of y or $\hat{y}$.

z: For the two-parameter Exponential distribution and three-parameter Gamma distribution, it represents the transformed variables: $x = y - y_o$ or $\hat{x} = \hat{y} - \hat{y}_o$. For the case of the three-parameter Lognormal distribution, it represents the natural log of x or $\hat{x}$.

β_z: Shape parameter for the two-parameters Exponential distribution and the three-parameter gamma distribution.

k_z: Scale parameter for the three-parameter Gamma distribution.

μ_z: Scale parameter of the three-parameter Log-normal distribution.

σ_z: Shape parameter for the three-parameter log-normal distribution.

Φ: The CDF of the standard normal distribution.

4.2.1 The one and two two-parameter Exponential distributions

If $Z = W - w_o$ has an exponential distribution with scale parameter β_z, and w_o *is known*, then the distribution of Z becomes a *one-parameter* Exponential distribution with parameter β_z.

The CDF of Z in this case is:

$$F(w) = \begin{cases} 1 - exp\left(-\frac{z}{\beta_Z}\right), w \geq w_0, \\ 0, \quad w < w_0 \end{cases} \qquad \text{Eq. 4.1}$$

with mean:

$$\mu_w = w_o + \beta_Z \qquad \text{Eq. 4.2}$$

variance:

$$\sigma_w^2 = \beta_Z^2 \qquad \text{Eq. 4.3}$$

and skewness coefficient:

$$\Upsilon_w = 2 \qquad \text{Eq. 4.4}$$

If $Z = W - w_o$ has an exponential distribution with scale parameter β_z, and w_o *is unknown*, then the distribution of W becomes a *two-parameter* Exponential distribution with parameters $\boldsymbol{\theta} = (w_o, \beta_z)$(Thomopoulos, 2017). The CDF of this distribution is obtained by replacing z by $w - w_o$ in Eq. 4.1.

Parameter estimation: Method of moments

If w_o is known (*one-parameter* Exponential distribution), β_z can be estimated from the sample mean by:

$$\hat{\beta}_z = \overline{w} - w_o \qquad \text{Eq. 4.5}$$

where $\overline{w}$ is the sample mean of W.

If w_o is unknown (*two-parameter* Exponential distributions), then β_z and w_o can be estimated from the mean and the standard deviation of the sample by:

$$\hat{\beta}_z = s_w \qquad \text{Eq. 4.6}$$

$$\widehat{w}_o = \overline{w} - s_w \qquad \text{Eq. 4.7}$$

where s_w is the sample standard deviation of W.

4.2.2 The two and three-parameter Lognormal distribution

If $Z = ln(W - w_o)$ is normally distributed with a mean μ_z and standard deviation σ_z, and w_o is known, then the distribution of Z becomes a two-parameter Lognormal distribution with parameters $\boldsymbol{\theta} = (\mu_Z, \sigma_Z)$ (Stedinger, 1980; Stedinger *et al.*, 1993; Thomopoulos, 2017).

The CDF of Z in this case is:

$$F(w) = \Phi\left(\frac{ln(z) - \mu_Z}{\sigma_Z}\right), w > w_0 \qquad \text{Eq. 4.8}$$

with mean:

$$\mu_w = w_o + exp(\mu_z + 0.5\sigma_z^2) \qquad \text{Eq. 4.9}$$

variance:

$$\sigma_w^2 = exp(\sigma_z^2 + 2\mu_z)[exp(\sigma_z^2) - 1] \qquad \text{Eq. 4.10}$$

and skewness coefficient:

$$\Upsilon_w = 3\phi + \phi^3 \qquad \text{Eq. 4.11}$$

where Φ is the CDF of the standardized normal distribution and $\phi = [exp(\sigma_z^2) - 1]^{0.5}$.

If $Z = ln(W - w_o)$ is normally distributed with mean μ_z and standard deviation σ_z, and w_o is unknown, then the distribution of W becomes a three-parameter Lognormal distribution with parameters $\boldsymbol{\theta} = (w_o, \mu_Z, \sigma_Z)$(Thomopoulos, 2017). The CDF of this distribution is obtained by replacing z by $w - w_o$ in Eq. 4.8.

Parameter estimation: Method of moments

If w_o is known (two-parameter Lognormal distribution), then μ_z and σ_z can be estimated from the mean and variance of the sample by:

$$\hat{\mu}_z = 2ln(\overline{w}) - 0.5ln(s_w^2 + \overline{w}^2) \qquad \text{Eq. 4.12}$$

$$\hat{\sigma}_z = \sqrt{-2ln(\overline{w}) + ln(s_w^2 + \overline{w}^2)} \qquad \text{Eq. 4.13}$$

where s_w^2 is the variance of the sample of W.

If w_o is unknown (three-parameter Lognormal distribution), then σ_z has to be first estimated from the sample estimate of the skewness coefficient $\hat{\Upsilon}_w$ based on the following relationship given by Chow (1954) and by Wilson & Worcester (1945).

$$exp(\hat{\sigma}_z^2) = \left\{1 + 0.5\left[\hat{\Upsilon}_w^2 + \hat{\Upsilon}_w\left(4 + \hat{\Upsilon}_w^2\right)^{0.5}\right]\right\}^{1/3} + \left\{1 + 0.5\left[\hat{\Upsilon}_w^2 - \hat{\Upsilon}_w\left(4 + \hat{\Upsilon}_w^2\right)^{0.5}\right]\right\}^{1/3} - 1 \quad \text{Eq. 4.14}$$

μ_z and w_o can be then estimated by:

$$\hat{\mu}_z = 0.5ln\left[\frac{s_w^2}{exp(\hat{\sigma}_z^2)[exp(\hat{\sigma}_z^2) - 1]}\right] \quad \text{Eq. 4.15}$$

$$\hat{w}_o = \overline{w} - exp(\hat{u}_z + 0.5\hat{\sigma}_z^2) \quad \text{Eq. 4.16}$$

4.2.3 The two and three-parameter Gamma distribution

If $Z = W - w_o$ is a random variable distributed according to a Gamma distribution with scale parameter β_z and shape parameter k_z, and w_o is known, then the distribution of Z becomes a two-parameter Gamma distribution with parameter $\boldsymbol{\theta} = (\beta_z, k_z)$ (Bowman and Shenton, 2011).

The CDF of Z is:

$$F(w) = \frac{1}{\Gamma(k_z)}\gamma\left(k_z, \frac{w - w_o}{\beta_z}\right), w > w_0 \quad \text{Eq. 4.17}$$

with mean:

$$\mu_w = w_o + k_z\beta_z \quad \text{Eq. 4.18}$$

variance:

$$\sigma_w^2 = k_z\beta_z^2 \quad \text{Eq. 4.19}$$

and skewness coefficient:

$$\Upsilon_W = \frac{2}{\sqrt{k_z}} \quad \text{Eq. 4.20}$$

Where Γ is the gamma function.

If $Z = W - w_o$ is a random variable distributed according to a Gamma distribution with scale parameter β_z and shape parameter k_z, and w_o is *unknown*, then the

distribution of Z becomes a three-parameter Gamma distribution with parameter $\boldsymbol{\theta} = (w_o\ , \beta_z, k_z)$

Parameter estimation: Method of moments

If w_o is known (two-parameter Gamma distribution), then β_z and k_z can be estimated from the mean and variance of the sample by:

$$\hat{k}_z = \frac{\overline{w} - w_o}{\hat{\beta}_z} \qquad \text{Eq. 4.21}$$

$$\hat{\beta}_z = \left(\frac{s_w^2}{\hat{k}_z}\right)^{0.5} \qquad \text{Eq. 4.22}$$

If w_o is unknown (three-parameter Gamma distribution), then k_z has to be first estimated from the sample estimate of the skewness coefficient $\hat{\Upsilon}_w$ by:

$$\hat{k}_z = \left(\frac{2}{\hat{\Upsilon}_w}\right)^2 \qquad \text{Eq. 4.23}$$

β_z can be then estimated by using Eq. 4.22. Finally, w_o can be estimated by:

$$\hat{w}_o = \overline{w} - \hat{\beta}_Z \hat{k}_Z \qquad \text{Eq. 4.24}$$

4.3 Bivariate distributions.

This section describes/explains the theory of some bivariate distributions, which are important to understand the criteria used in this work to do the bivariate modelling of y and $\hat{y}$.

4.3.1 The standardized bivariate Normal distribution (BND).

The bivariate normal distribution (BND) in its standard form is usually the basis, and this work is not an exception, of algorithms used to simulate any bivariate distribution. The BND is not in itself a suitable model of flood peaks and their forecasts, but it underpins models that can be, for example, the Lognormal. Therefore, this section of the chapter describes the properties of this bivariate distribution in the standardized form with $N(0,1)$ variables which are important to understand the algorithms which can be used to simulate y and $\hat{y}$.

Univariate properties

Both marginal distributions are Normal. Thus, suppose that H and $\widehat{\mathrm{H}}$ are correlated standardized Normal random variables; the corresponding CDFs of these marginals are defined by:

$$F(\eta) = \frac{1}{2} erf\left(\frac{\eta}{\sqrt{2}}\right) \quad \text{Eq. 4.25}$$

$$F(\hat{\eta}) = \frac{1}{2} erf\left(\frac{\hat{\eta}}{\sqrt{2}}\right) \quad \text{Eq. 4.26}$$

where erf is a function called the error function.

Bivariate properties

The joint probability density function (PDF) is:

$$f(\eta, \hat{\eta}) = \frac{1}{2\pi\sqrt{1-\rho_{\eta\hat{\eta}}^2}} exp\left[-\frac{1}{2\left(1-\rho_{\eta\hat{\eta}}^2\right)}\left(\hat{\eta}^2 + \eta^2 - 2\eta\hat{\eta}\rho_{\eta\hat{\eta}}\right)\right] \quad \text{Eq. 4.27}$$

where $\rho_{\eta\hat{\eta}}$ is the product-moment correlation coefficient between the two standardized variables.

Conditional properties

The conditional density of $(\eta|\hat{\eta})$ is Normal and, for the case of H given $\widehat{\mathrm{H}} = \hat{\eta}$, has the following characteristics:

$$f(\eta|\hat{\eta}) = \frac{1}{2\pi\sqrt{1-\rho_{\eta\hat{\eta}}^2}} exp\left[-\frac{1}{2\left(1-\rho_{\eta\hat{\eta}}^2\right)}\left(\rho_{\eta\hat{\eta}}^2\hat{\eta}^2 + \eta^2 - 2\eta\hat{\eta}\rho_{\eta\hat{\eta}}\right)\right] \quad \text{Eq. 4.28}$$

with conditional mean:

$$E\left(\mathrm{H}\middle|\widehat{\mathrm{H}} = \hat{\eta}\right) = \mu_{\eta|\hat{\eta}} = \hat{\eta}\rho_{\eta\hat{\eta}} \quad \text{Eq. 4.29}$$

and conditional variance:

$$var\left(\mathrm{H}\middle|\widehat{\mathrm{H}}=\hat{\eta}\right)=\sigma_{\eta|\hat{\eta}}^{2}=1-\rho_{\eta\hat{\eta}}^{2} \qquad \text{Eq. 4.30}$$

The mean regression of H on $\widehat{\mathrm{H}}$, i.e., the regression of the values defined by $E\left(\mathrm{H}\middle|\widehat{\mathrm{H}}=\hat{\eta}\right)$, is linear, the conditional variance is constant, and the probability distribution of the estimation errors, namely $\varepsilon=\eta-\mu_{\eta|\hat{\eta}}$ is a Normal distribution with mean $\mu_{\varepsilon}=0$ and variance $\sigma_{\varepsilon}^{2}=1-\rho_{\eta\hat{\eta}}^{2}$.

4.3.2 The bivariate Exponential distribution (BED).

If y and $\hat{y}$ can be represented by the Exponential distribution, then the BED can be a potential candidate to conduct the bivariate modelling of these variates. The properties of this distribution can be derived from the Moran–Downton BED (Downton, 1970). An extensive discussion of this BED is found in Nagao & Kadoya (1971). Thus, this section first describes the standardized BED, which is the basis for describing the three- and five-parameters BED.

4.3.2.1 The standardized BED

The following considerations are for a BED distribution function in which every variable is standardized.

Univariate properties

Both marginal distributions are exponential. Thus, suppose that V and $\hat{V}$ are correlated standardized exponential variables; the corresponding CDFs of these marginals are defined by:

$$F(v)=1-e^{-v} \qquad \text{Eq. 4.31}$$

$$F(\hat{v})=1-e^{-\hat{v}} \qquad \text{Eq. 4.32}$$

Bivariate properties

The joint PDF is:

$$f(v,\hat{v})=\frac{1}{(1-\rho_{v\hat{v}})}exp\left[-\frac{v+\hat{v}}{1-\rho_{v\hat{v}}}\right]I_{o}\left(\frac{2\sqrt{v\hat{v}\rho_{v\hat{v}}}}{1-\rho_{v\hat{v}}}\right) \qquad \text{Eq. 4.33}$$

where I_o is the modified Bessel function of the first kind of order zero, and $\rho_{v\hat{v}}$ is the product-moment correlation coefficient between the two random variates (Balakrishna and Lai, 2009).

Expressed as an infinite series, the joint CDF is (Balakrishna and Lai, 2009):

$$F(v,\hat{v}) = (1-e^{-v})\left(1-e^{\hat{v}}\right) + \sum_{j=0}^{\infty} \frac{\rho_{v\hat{v}}^{j+1}}{(j+1)^2} L_j^{(1)}(v)\, L_j^{(1)}(\hat{v}) v\hat{v} e^{-(v+\hat{v})} \quad \text{Eq. 4.34}$$

where the $L_j^{(1)}$ are Laguerre polynomials.

Conditional properties

The conditional PDF, for the case of V given $\hat{V} = \hat{v}$, has the following characteristics:

$$f(v|\hat{v}) = \frac{1}{(1-\rho_{v\hat{v}})} exp\left[-\frac{v+\hat{v}\rho_{v\hat{v}}}{1-\rho_{v\hat{v}}}\right] I_o\left(\frac{2\sqrt{v\hat{v}\rho_{v\hat{v}}}}{1-\rho_{v\hat{v}}}\right) \quad \text{Eq. 4.35}$$

with conditional mean:

$$E\left(V\middle|\hat{V}=\hat{v}\right) = \mu_{v|\hat{v}} = 1 + \rho_{v\hat{v}}(\hat{v}-1) \quad \text{Eq. 4.36}$$

and conditional variance:

$$Var\left(V\middle|\hat{V}=\hat{v}\right) = \sigma_{v|\hat{v}}^2 = (1-\rho_{v\hat{v}})^2 + 2\hat{v}\rho_{v\hat{v}}(1-\rho_{v\hat{v}}) \quad \text{Eq. 4.37}$$

The mean regression of V on $\hat{V}$, i.e., the regression of the values defined by $E\left(V\middle|\hat{V}=\hat{v}\right)$, is linear with varying dispersion. The latter is based on the fact that, unlike the Normal distribution, the conditional variance is a function of the independent variable, in this case, denoted by $\hat{v}$. That means that, in the BED, the conditional variance increases as the value of the independent variable increases. This is an important property for modelling the conditional variance of an observed peak water value, given the forecast, as it has been observed to increase with the forecast value (Coccia and Todini, 2011; Zhao *et al.*, 2015).

4.3.2.2 The three-parameters BED

If y and $\hat{y}$ are correlated and y_o and $\hat{y}_o$ are known a priori, then the transformed variables:

$$x = y - y_o \qquad \text{Eq. 4.38}$$

$$\hat{x} = \hat{y} - \hat{y}_o \qquad \text{Eq. 4.39}$$

can be represented by a one-parameter Exponential distribution with parameters β_x and $\beta_{\hat{x}}$, respectively (see section 4.2.1). These variables are usually known as threshold exceedances (Claps and Laio, 2003; Bezak *et al.*, 2014), and they will be referred to as such here. Pairs $(x, \hat{x})$ can be, therefore, represented by a three-parameter BED with parameters $\boldsymbol{\theta}_{x\hat{x}} = (\beta_x, \beta_{\hat{x}}, \rho_{x\hat{x}})$, where $\rho_{x\hat{x}}$ is the product-moment correlation between x and $\hat{x}$. The designation of these variables is $(x, \hat{x}) \sim BED(\beta_x, \beta_{\hat{x}}, \rho_{x\hat{x}})$.

For this BED, the analytical expression of $f(x, \hat{x})$ and $f(x|\hat{x})$ can be easily obtained from the analytical expressions of the standardized BED (Eqs. 4.36 and 4.37) by doing the following replacement:

$$v = \frac{x}{\beta_X} \qquad \text{Eq. 4.40}$$

$$\hat{v} = \frac{\hat{x}}{\beta_{\hat{X}}} \qquad \text{Eq. 4.41}$$

$$\rho_{v\hat{v}} = \rho_{x\hat{x}} \qquad \text{Eq. 4.42}$$

and adding the product $\beta_x \beta_{\hat{x}}$ and β_x to the denominators of the factors that represent the normalizing constant in $f(x, \hat{x})$ and $f(x|\hat{x})$, respectively. The expression of $F(x, \hat{x})$ is not needed as $F(v, \hat{v}) = F(x, \hat{x})$. The conditional moments are defined as:

$$E\left(X \middle| \hat{X} = \hat{x}\right) = \mu_{x|\hat{x}} = \beta_X \left(1 - \rho + \frac{\rho \hat{x}}{\beta_{\hat{X}}}\right) \qquad \text{Eq. 4.43}$$

$$Var\left(X \middle| \hat{X} = \hat{x}\right) = \sigma^2_{x|\hat{x}} = \beta_X \left[(1-\rho)^2 + \frac{2}{\beta_{\hat{X}}} \rho (1-\rho) \hat{x}\right] \qquad \text{Eq. 4.44}$$

4.3.2.3 The five-parameter BED

If y and $\hat{y}$ can be represented by the two-parameter Exponential distribution with parameters $\boldsymbol{\theta}_y = (y_o, \beta_x)$ and $\boldsymbol{\theta}_{\hat{y}} = (\hat{y}_o, \beta_{\hat{x}})$ respectively and they are correlated, the direct bivariate modelling of y and $\hat{y}$ should be based on the five-parameter BED with parameters $\boldsymbol{\theta}_{y\hat{y}} = \left(\beta_x, \beta_{\hat{x}}, \rho_{y\hat{y}}, y_o, \hat{y}_o\right)$, where $\rho_{y\hat{y}}$ is the product-moment correlation between y and $\hat{y}$. The designation of these variables is $(y, \hat{y}) \sim BED\left(\beta_x, \beta_{\hat{x}}, \rho_{y\hat{y}}, y_o, \hat{y}_o\right)$. The standardized BED can also be used to obtain the analytical expressions for this distribution. In this case, to obtain $f(y, \hat{y})$ and $f(y|\hat{y})$ the replacements in Eq. 4.33 and 4.35 should be:

$$v = \frac{y - y_o}{\beta_x} \qquad \text{Eq. 4.45}$$

$$\hat{v} = \frac{\hat{y} - \hat{y}_o}{\beta_{\hat{x}}} \qquad \text{Eq. 4.46}$$

$$\rho_{v\hat{v}} = \rho_{y\hat{y}} \qquad \text{Eq. 4.47}$$

Since the bivariate modelling of the pairs $(y, \hat{y})$ can be based on the conditional moments of the standardized exponential variables, this work did not derive the equations of the conditional moments $\mu_{y|\hat{y}}$ and $\sigma^2_{y|\hat{y}}$ which should include the location and scale parameters and which can be obtained from Eq 4.36 and Eq 4.37, respectively.

The five-parameter BED is convenient since it allows us to explore the case when the means of y and $\hat{y}$ are the same but their variances are different. This analysis could not be done if they would be represented by the three-parameter BED where the mean and standard deviation are equal.

4.3.3 The bivariate Log-normal distribution (BLND)

If y and $\hat{y}$ can be represented by univariate Log-normal distributions, then the BLND can be a potential candidate to conduct the bivariate modelling of these variates. A description of this bivariate distribution is found in Yue (2000) and Thomopoulos (2017). This section first describes the five-parameter BLND, which is the basis for describing the seven-parameter BLND.

4.3.3.1 The five-parameters BLND

If y and $\hat{y}$ are correlated and y_o and $\hat{y}_o$ are known a priori, then the natural log of their thresholds exceedances, i.e., $z = ln(x)$ and $\hat{z} = ln(\hat{x})$, are normally distributed with means μ_z and $\mu_{\hat{z}}$ and standard deviations σ_z and $\sigma_{\hat{z}}$, respectively (see section 4.2.2). Therefore, the pairs $(x, \hat{x})$ can be represented by a five-parameter BLND with parameters $\boldsymbol{\theta}_{x\hat{x}} = (\mu_z, \sigma_z, \mu_{\hat{z}}, \sigma_{\hat{z}}, \rho_{z\hat{z}})$, where $\rho_{z\hat{z}}$ is the product-moment correlation coefficient between z and $\hat{z}$. The designation of these variables is $(x,\hat{x}) \sim BLND(\mu_z, \sigma_z, \mu_{\hat{z}}, \sigma_{\hat{z}}, \rho_{z\hat{z}})$.

Univariate properties

Both marginal distributions are lognormally distributed. Assuming that the pairs $(x, \hat{x})$ have these distributions; the corresponding CDFs are defined as:

$$F(x) = \Phi\left(\frac{ln(x) - \mu_Z}{\sigma_Z}\right), x > 0 \qquad \text{Eq. 4.48}$$

$$F(\hat{x}) = \Phi\left(\frac{ln(\hat{x}) - \mu_{\hat{Z}}}{\sigma_{\hat{Z}}}\right), \hat{x} > 0 \qquad \text{Eq. 4.49}$$

Bivariate properties

The joint PDF is:

$$f(x, \hat{x}) = \frac{1}{2\pi x \hat{x} \sigma_Z \sigma_{\hat{Z}} \sqrt{1 - \rho_{Z\hat{Z}}^2}} exp\left(-\frac{q}{2}\right)$$

where

Eq. 4.50

$$q = \frac{1}{1-\rho_{Z\hat{Z}}^2}\left[\left(\frac{ln(x)-\mu_Z}{\sigma_Z}\right)^2 - 2\rho_{Z\hat{Z}}\left(\frac{ln(x)-\mu_Z}{\sigma_Z}\right)\left(\frac{ln(\hat{x})-\mu_{\hat{Z}}}{\sigma_{\hat{Z}}}\right) + \left(\frac{ln(\hat{x})-\mu_{\hat{Z}}}{\sigma_{\hat{Z}}}\right)^2\right]$$

Conditional properties

The conditional PDF is lognormally distributed and, for the case of X given $\hat{X} = \hat{x}$, has the following characteristics:

$$f(x|\hat{x}) = \frac{1}{x\sigma_{z|\hat{z}}\sqrt{2\pi}} exp\left[-\frac{1}{2}\left(\frac{ln(x) - \mu_{z|\hat{z}}}{\sigma_{z|\hat{z}}}\right)^{0.5}\right] \quad \text{Eq. 4.51}$$

with conditional mean:

$$\mu_{x|\hat{x}} = \mu_z + \rho\frac{\sigma_z}{\sigma_{\hat{z}}}[ln(\hat{x}) - \mu_{\hat{z}}] \quad \text{Eq. 4.52}$$

and conditional variance:

$$\sigma_{x|\hat{x}} = \sigma_z\sqrt{1-\rho^2} \quad \text{Eq. 4.53}$$

Note that, for the sake of simplicity, the properties of BLND $(\mu_z, \sigma_z, \mu_{\hat{z}}, \sigma_{\hat{z}}, \rho_{z\hat{z}})$ characterising the pairs $(x, \hat{x})$ have been described by taking the natural log of each of the two variables where the Normal distribution emerges, and this distribution is easier to handle. Thus, in fact, the above equations represent the BND of the variables $z = ln(x)$ and $\hat{z} = ln(\hat{x})$. Therefore, the characteristics of the mean regression and conditional variance are the same as those described in section 4.3.1. Some further properties of the BLND $(\mu_Z, \sigma_Z, \mu_{\hat{Z}}, \sigma_{\hat{Z}}, \rho_{Z\hat{Z}})$ in terms of the log-normal variables, in this case, denoted as x and $\hat{x}$, can be found in Balakrishna and Lai (2009). It should be noted that, while the conditional variance in the Normal space is constant, this is not the case in the logNormal space where it is a function of x as a result of the logarithmic transformation (Eq 4.53).

4.3.3.2 The seven-parameters BLND.

If y and $\hat{y}$ can be represented by a three-parameter Log-Normal distribution with parameters $\boldsymbol{\theta_y} = (y_o, \mu_Z, \sigma_Z)$ and $\boldsymbol{\theta_{\hat{y}}} = (\hat{y}_o, \mu_{\hat{Z}}, \sigma_{\hat{Z}})$ respectively, and they are correlated, the bivariate modelling of y and $\hat{y}$ should be based on the seven-parameters BLND with parameters $\boldsymbol{\theta_{y,\hat{y}}} = (\mu_z, \sigma_z, \mu_{\hat{z}}, \sigma_{\hat{z}}, y_o, \hat{y}_o, \rho_{z\hat{z}})$. The designation of these variables is $(y,\hat{y}) \sim BLND(\mu_z, \sigma_z, \mu_{\hat{z}}, \sigma_{\hat{z}}, y_o, \hat{y}_o, \rho_{z\hat{z}})$. For this BLND, the analytical expression of univariate and bivariate properties can be easily obtained from the analytical expressions of the five-parameter BLND by substituting $x = y - y_o$ and $\hat{x} = \hat{y} - \hat{y}_o$, respectively (see the previous section).

4.4 Bivariate simulation.

The simulation of any bivariate distribution can be tackled by different approaches (Balakrishna & Lai, 2009; Lewis et al., 2017). In this research, the bivariate simulation of y and $\hat{y}$ was based on the method known as the conditional approach (Lewis and Orav, 2018) by using algorithms that have no restrictions on the shapes of the marginal distributions. The Gaussian copula was employed for these purposes. This section thus starts by explaining the conditional approach and how it is related to the concept of PU, then describes the Gaussian copula approach, and finally shows the algorithms built to simulate pairs $(y,\hat{y})$ and to build the PU of $y|\hat{y}$ assuming they can be represented by a BED or a BLND. By way of example, for the case when the pairs $(y,\hat{y}) \sim BED(\beta_x, \beta_x, y_o, \hat{y}_o, \rho_{yy})$, an algorithm that considers restrictions on the shape of the marginal distribution was considered. This algorithm is based on the Moran-Downton BED, and an analysis of its advantages and disadvantages with respect to the Gaussian-copula-based algorithm is provided.

4.4.1 The conditional simulation approach and predictive uncertainty (PU)

Since PU can be defined as the uncertainty of a future realization of a predictand (the future quantity of interest) conditional on, for example, one model forecast (Todini, 2016), the conditional distribution of y given $\hat{y}_i$, i.e. $f(y|\hat{y}_i)$, can be used to derive PU. This concept can be further used to represent the realization of $\hat{y}_i$, i.e. y_i, through a value drawn from $f(y|\hat{y}_i)$. This approach is known in bivariate simulation as the conditional approach (Lewis and Orav, 2018). When using this approach, the ideal situation is when one has the moment equations of $f(y|\hat{y}_i)$. If these equations are not available or are not easy to handle, $f(y|\hat{y}_i)$ can be built via a sampling technique where the conditional approach is applied in a different space, and, then, these values are converted to the real space by using a transformation technique. In the next subsections, it is shown how this approach can be used to do the bivariate simulation of y and $\hat{y}$ assuming different bivariate probabilistic models (BED and BLND) through a Gaussian copula-based approach.

4.4.2 Bivariate simulation based on the Gaussian copula

The most convenient approaches to do a bivariate simulation are those that do not impose any restriction on the range of association between variates or with the form of the marginal distributions (Kelly and Krzysztofowicz, 1997), i.e., marginal-free

bivariate generators. In this work, one of these methods, the Gaussian copula, was employed to do the bivariate simulation of y and $\hat{y}$.

The Gaussian copula, in the bivariate version, is, in essence, the bivariate CDF derived from the standard bivariate Normal distribution (BND), whose univariate marginal CDFs are uniformly distributed on the range [0,1]. Thus, this method basically uses the standard BND to simulate bivariate variables with a specific correlation structure and then transform the standardized normal distributions into uniform distributions through the probability integral transform (PIT). Finally, the uniform distributions are transformed into any required distributions by computing the inverses of the required distributions associated with the values of the uniform distributions.

The Gaussian copula approach must first generate jointly standardized Normal random variables, i.e., pairs $(\eta, \hat{\eta})$ with a specified correlation $\rho_{\eta\hat{\eta}}$ (according to the notation used in section 4.3.1). Based on these bivariate points, a generic Gaussian copula generator is constructed as follows. Let (S, Q) denote a vector of continuous random variables with arbitrarily specified, strictly increasing, and continuous probability distributions. Then, the bivariate simulation of the joint distribution of the bivariate pairs (s, q) can be done by the following approach.

- ✓ **Step 1:** Generate a set of n random values of $\hat{H}$, i.e.$\hat{\eta}$, from $N\sim(0,1)$.
- ✓ **Step 2:** Generate η_i from $f(n|\hat{n}_i)$ (see Eq. 4.28), i.e., the conditional distribution of H given the value $\hat{\eta}_i$ of $\hat{H}$ generated in Step 1, by drawing a random value from each of these conditional distributions.
- ✓ **Step 4:** Compute the CDFs of each of the corresponding pairs $(\eta, \hat{\eta})$, denoted as $F(\eta)$ and $F(\hat{\eta})$; these then define Uniformly distributed pairs of random variables as
  $u = F(\eta)$ and $r = F(\hat{\eta})$
- ✓ **Step 5:** Using the probability integral transform (PIT), obtain s and q as:
  $s = F_S^{-1}(u)$ and $q = F_Q^{-1}(r)$
  where F_S^{-1} and F_Q^{-1} are the inverses of the CDFs of s and q, respectively.

Figure 4.2 illustrates the bivariate simulation of pairs $(\eta, \hat{\eta})$ through Steps 1 and 2 of the algorithm mentioned above.

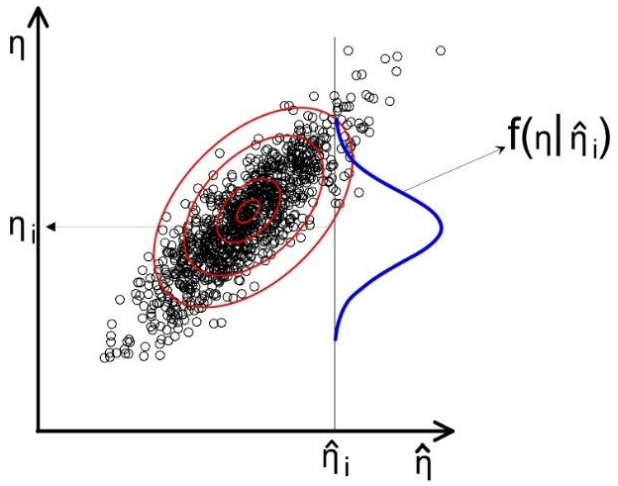


Figure 4.2: Illustration of how pairs (η,η̂) are generated from a standard BND based on the conditional approach

4.4.3 Bivariate simulation of y and $\hat{y}$ based on the BED

The bivariate simulation of y and $\hat{y}$ based on the BED was addressed by considering two algorithms: i) an algorithm based on the Moran–Downton BED; and ii) an algorithm based on the Gaussian copula. The concepts, procedures, and advantages and disadvantages of each of them are explained as follows.

4.4.3.1 Simulation of y and $\hat{y}$ based on the Moran-Downton BED

The Moran-Downton BED algorithm is based on the generation of standardized bivariate Exponential random variables; thus, this section first explains how it is done based on the conditional approach and then describes the algorithms used to simulate pairs (y,$\hat{y}$) and to derive the PU.

Using the notation of section 4.3.2, pairs $(v, \hat{v})$ of a standard BED with a specified correlation $\rho_{v\hat{v}}$ can be simulated based on the conditional approach by:

- ✓ **Step 1:** Generate a random value of $\hat{V}$, i.e. $\hat{v}_i$, from $Exp{\sim}(1)$ i.e., standardised exponential distribution with a mean 1.
- ✓ **Step 2:** Generate v_i from $f(v|\hat{v}_i)$ (see Eq. 4.35), i.e., the conditional distribution of V given the value $\hat{v}_i$ of $\hat{V}$ generated in Step 1, by drawing a random value from this conditional distribution.

Assuming $(y,\hat{y}) \sim BED\left(\beta_x, \beta_x, y_o, \hat{y}_o, \rho_{yy}\right)$, pairs $(y,\hat{y})$ can be simulated from the Moran-Downton BED by following the steps:

- ✓ **Step 1:** Define the moments β_x and $\beta_{\hat{x}}$, the coefficient of correlation $\rho_{y\hat{y}}$, and the thresholds y_o and $\hat{y}_o$.

- ✓ **Step 2:** Use the relationship $\rho_{y\hat{y}} = \rho_{v\hat{v}}$ to generate pairs $(v, \hat{v})$ based on the conditional approach (see Figure 4.3a).
- ✓ **Step 3:** Obtain the exceedances of y_o and $\hat{y}_o$ i.e., the pairs $(x,\hat{x})$, by:
  $x = \beta_x v$
  $\hat{x} = \beta_{\hat{x}} \hat{v}$
- ✓ **Step 4:** Obtain the values of y and $\hat{y}$ by adding the thresholds to the exceedances:
  $y = y_o + x$
  $\hat{y} = \hat{y}_o + \hat{x}$

Figure 4.3 illustrates the above algorithm.

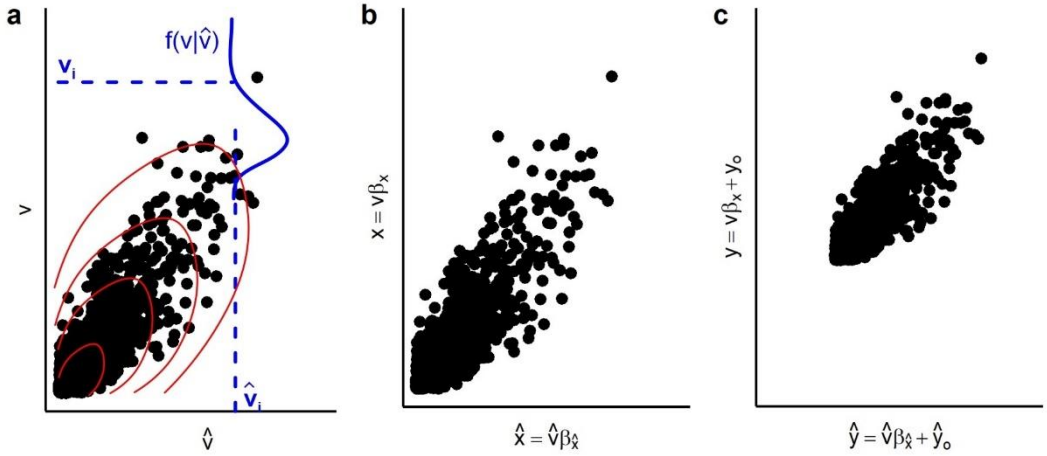


Figure 4.3: Illustration of the Moran-Downton model-based algorithm used to generate pairs $(y,\hat{y})$

The algorithm starts generating pairs (v,v) based on the conditional approach (Figure a); then the pairs (x,x) are obtained by multiplying v and $\hat{v}$ by the scale parameters β_x and $\beta_{\hat{x}}$ respectively (Figure b); Finally, the pairs $(y,\hat{y})$ are obtained by adding to each value of x and $\hat{x}$ a constant value defined by the thresholds y_o and $\hat{y}_o$, respectively (Figure c). In this algorithm, the following relationship is met: $\rho_{v\hat{v}} = \rho_{x\hat{x}} = \rho_{y\hat{y}}$.

Building of PU

The PU can be built by simply sampling n values from the conditional distribution $f(v|\hat{v}_i)$ obtained in the algorithm used to obtain the pairs $(y,\hat{y})$ (Step 2) and then conveying the resulting values to real space by following the steps of the algorithm in terms of y. PU can be finally expressed in terms of density values by, for example, computing the Kernel density estimations (Sheather and Jones, 1991) of the sampling values.

4.4.3.2 Bivariate simulation of y and $\hat{y}$ based on the Gaussian copula

Assuming $(y,\hat{y}) \sim BED(\beta_x, \beta_x, y_o, \hat{y}_o, \rho_{yy})$, pairs $(y,\hat{y})$ can be simulated based on the Gaussian copula through the following algorithm.

- ✓ **Step 1**: Define the moments β_x and $\beta_{\hat{x}}$, the coefficient of correlation $\rho_{n\hat{n}}$ associated with $\rho_{y\hat{y}}$, and the thresholds y_o and $\hat{y}_o$.
- ✓ **Step 2**: Generate pairs $(\eta, \hat{\eta})$ based on the conditional approach (Figure 4.2).
- ✓ **Step 3**: Compute the CDF of the corresponding pairs $(\eta, \hat{\eta})$.
  $u = F(\eta)$ and $r = F(\hat{\eta})$
- ✓ **Step 4**: Compute standardized exponential variables v and $\hat{v}$ as:
  $v = F_V^{-1}(u)$ and $\hat{v} = F_{\hat{V}}^{-1}(r)$
  where F_V^{-1} and $F_{\hat{V}}^{-1}$ are the inverse of the distributions of v and $\hat{v}$ which are in turn defined as:
  $v = \frac{y - y_o}{\beta_x}$ and $\hat{v} = \frac{\hat{y} - \hat{y}_o}{\beta_{\hat{x}}}$
- ✓ **Step 5**: Pairs $(y, \hat{y})$ can be finally obtained as:
  $y = y_o + \beta_x v$
  $\hat{y} = \hat{y}_o + \beta_{\hat{x}} \hat{v}$

Building the PU

The PU can be built by simply sampling n values from the conditional distribution $f(\eta|\hat{\eta}_i)$ obtained in the algorithm used to obtain the pairs $(y,\hat{y})$ (Step 2) and then converting the resulting values to real space by following the steps of the algorithm in terms of y. PU can be finally expressed in terms of density values by computing the Kernel density estimations of the sampling values.

4.4.3.3 Advantages and disadvantages of Moran-Downton-model- and Gaussian copula-based algorithm

The advantages and disadvantages of the algorithms used to simulate pairs (y,y) based on the BED are listed below.

- ✓ The Moran–Downton model-based algorithm uses the value of the correlation coefficient in the real space, i.e., $\rho_{y\hat{y}}$. It can be considered an advantage with respect to the Gaussian copula-based approach as the latter bases the bivariate modelling on the normal space, for which they must use a transformation technique to obtain the value of the correlation coefficient in the real space. This could be an issue if one does not have an analytical relationship between these two variables.
- ✓ Both algorithms have no restriction with the range of association between y and $\hat{y}$.
- ✓ The Moran–Downton model has a restriction with the shape of the marginals; they must be exponentially distributed, i.e., it is a non-marginal free bivariate generator. The Gaussian copula is a marginal-free bivariate generator.
- ✓ The Moran–Downton-model- based algorithm allows the theoretical joint probabilities to be computed (see Eq. 4.34). It can be considered an advantage in terms of the Bivariate Goodness of fit test (GoF) since one can compare them with the resulting empirical joint probabilities obtained from the standardized variables of a sample of pairs $(y,\hat{y})$.

4.4.4 Bivariate Simulation of y and $\hat{y}$ based on the seven-parameters BLND

Since there is a close relationship between Normal and Lognormal distributions, several analytical expression has been developed to show that relationship (Lai et al., 1999; Thomopoulos, 2017). In this work, some of these expressions have been used to build an algorithm that allows us to simulate pairs $(y,\hat{y}) \sim BLND(\mu_z, \sigma_z, \mu_{\hat{z}}, \sigma_{\hat{z}}, y_o, \hat{y}_o, \rho_{z\hat{z}})$ (see section 4.3.3.2) and build the PU. This algorithm is detailed below.

Considering the notation used in section 4.3.3 and assuming $(y,\hat{y}) \sim BLND(\mu_z, \sigma_z, \mu_{\hat{z}}, \sigma_{\hat{z}}, y_o, \hat{y}_o, \rho_{z\hat{z}})$, pairs $(y,\hat{y})$ can be simulated through the following algorithm.

- ✓ **Step 1:** Define the moments μ_y, $\mu_{\hat{y}}$, σ_y, $\sigma_{\hat{y}}$, the coefficient of correlation $\rho_{y\hat{y}}$, and the thresholds y_o and $\hat{y}_o$.

- ✓ **Step 2**: Use the expression $\mu_x = \mu_y - y_o$, $\mu_{\hat{x}} = \mu_{\hat{y}} - \hat{y}_o$, $\sigma_x^2 = \sigma_y^2$, and $\sigma_{\hat{x}}^2 = \sigma_{\hat{y}}^2$ to obtain the mean and variance of the log-transformed values of y and $\hat{y}$ in the Normal space by using the following analytical relationships between the Normal and log-Normal distribution:
  $\mu_z = 2\ln(\mu_x) - 0.5\ln(\mu_x^2 + \sigma_x^2)$
  $\sigma_z^2 = -2\ln(\mu_x) + \ln(\mu_x^2 + \sigma_x^2)$
  $\mu_{\hat{z}} = 2\ln(\mu_{\hat{x}}) - 0.5\ln(\mu_{\hat{x}}^2 + \sigma_{\hat{x}}^2)$
  $\sigma_{\hat{z}}^2 = -2\ln(\mu_{\hat{x}}) + \ln(\mu_{\hat{x}}^2 + \sigma_{\hat{x}}^2)$

- ✓ **Step 3:** Use the expression $\rho_{x\hat{x}} = \rho_{y\hat{y}}$ to compute the coefficient of correlation of z=ln(x) and $\hat{z}$ = ln($\hat{x}$) in the normal space by:
  $$\rho_{z\hat{z}} = \frac{\ln\left[\left(\rho_{y\hat{y}} * \sqrt{(\exp(\sigma_z^2) - 1) * (\exp(\sigma_{\hat{z}}^2) - 1)}\right) - 1\right]}{\sigma_z * \sigma_{\hat{z}}}$$

- ✓ **Step 4:** Use the relationship $\rho_{z\hat{z}} = \rho_{\eta\hat{\eta}}$ to generate pairs $(\eta, \hat{\eta})$ based on the conditional approach (see Figure 4.2).

- ✓ **Step 5**: Obtain the values of the transformed variables of y and $\hat{y}$ in the normal space, i.e., the pairs $(z,\hat{z})$, by:
  $z = \mu_z + \sigma_z * \eta$
  $\hat{z} = \mu_{\hat{z}} + \sigma_{\hat{z}} * \hat{\eta}$

- ✓ **Step 6**: Obtain the pairs $(y,\hat{y})$ by.
  $y = \exp(z) + y_o$
  $\hat{y} = \exp(\hat{z}) + \hat{y}_o$

Building the PU

The PU of each pair $(y_i,\hat{y}_i)$ can be built by simply sampling n values from the conditional distribution $f(\eta|\hat{\eta}_i)$ obtained in the algorithm used to obtain the pairs $(y,\hat{y})$ (Step 4) and then conveying the resulting values to real space by following the steps of the algorithm in terms of y. PU can be finally expressed in terms of density values by computing the Kernel density estimations of the sampling values.

4.5 Data analysis.

Data analysis was conducted to support the assumptions made when doing the bivariate modelling of pairs $(y,\hat{y})$, which basically consisted of doing a univariate and bivariate analysis of these variates.

4.5.1 Observed data (y)

The analysis of y was based on the records of four gauging stations located close to or in a floodplain in England (see Figure 4.4). These stations were selected because they are at locations with known flooding. The observed discharges were converted into water levels y by using the rating curves of the at-site gauges taken from the National River Flow Archive (NRFA, see https://nrfa.ceh.ac.uk/).

Figure 4.4: Geographical location of the gauging stations used to analyse y and $\hat{y}$

Table 4.2 shows general information for the stations used in the univariate and bivariate analysis of y.

Table 4.2 Description of the available information used to analyse y and $\hat{y}$.

NRFA Code	Name	Lat	Long	Record length (15-min time step)	Period simulated
24003	River Wear at Stanhope	54.75	-2.03	Jan-1961/ Apr-2014	-
23006	River South Tyne at Featherstone	54.94	-2.51	Oct-1967/ Apr-2014	-
22007	River Wansbeck at Mitford	55.17	-1.73	Mar-1963/ Jan-2015	-
76007	River Eden at Sheepmount	54.90	-2.95	Jan-1976/ Nov-2012	Jan-1990/ Dec-2006

4.5.2 Simulated forecast data ($\hat{y}$)

The analysis of $\hat{y}$ was done for only one of these gauging stations, the River Eden at Sheepmount, for which a previously calibrated SHETRAN model was available. The main criterion to be satisfied was that a sufficiently long record was available to calibrate and validate the model, to support the analysis of y and $\hat{y}$, and the Eden SHETRAN model met this criterion. Setting up and calibrating models for other stations would have taken a disproportionate amount of time and was not the main focus of the study.

SHETRAN (Ewen *et al.*, 2000) is a state-of-the-art physically-based, spatially distributed model that has been developed at Newcastle University and applied widely to catchments in the UK and across the world (Lewis *et al.*, 2018). SHETRAN uses a grid-based representation of the catchment, and inputs are often gridded at a resolution between 50 m and 5km. The timestep used is often hourly or daily, although this temporal resolution decreases during storm events to better represent rapid infiltration and surface runoff processes. It has been used in countless applications in hydrology, such as in the analysis of the impact of groundwater abstractions on streamflows (Parkin *et al.*, 2007), deforestation impacts on peak flows and sediments yields (Birkinshaw *et al.*, 2011), nitrate transport (Koo and O'Connell, 2006), and real-time flood forecasting (Mellor *et al.*, 2000) to name but a few.

Detailed information on the SHETRAN model used to derive $\hat{y}$ (the Eden SHETRAN model) can be found in Janes *et al.* (2018). It was built to study the bank erosion process in the Eden catchment, and its calibrated hydrological component

was used in this research to represent simulated discharges of the River Eden at Sheepmount (Figure 4.5). The Eden SHETRAN model was applied to a 2400 km^2 predominately rural Eden catchment. A 1-km^2 grid resolution and a timestep of 1 hour were considered. A GIS methodology was used to set up the model based on a 30 m digital elevation model, land-use, and soil maps. The spatial hourly rainfall was derived by disaggregating a daily 1 km^2 gridded daily rainfall product from 1990–2007. The 1991-2001 period was used for the calibration procedure, whereas the 2001-2007 period for validation (with 1-year 'start-up' period). The calibration and validation procedure considered hourly and daily data from the NRFA gauging stations and HiFlows data sets. The hourly hydrological performance of the model based on the data of the River Eden at Sheepmount was characterised by a value of the Nash-Sutcliffe Efficiency (NSE) of 0.90, and a coefficient of determination (R^2) of 0.91. Figure 4.6 shows the simulated and observed hourly discharges for the validation period.

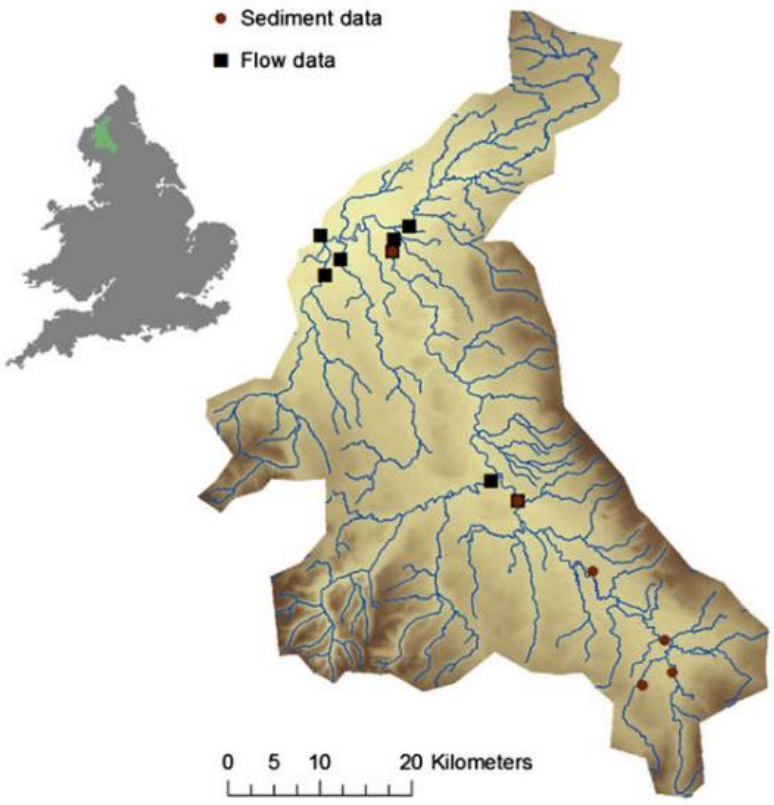


Figure 4.5: Eden catchment, Cumbria UK.
Locations of gauging stations (black) and sediment data (brown) used for the calibration and validation of the model are shown (taken from Janes *et al.*(2018)).

The SHETRAN model was developed for simulation, not real-time forecasting, and so no data assimilation/ forecast updating was undertaken. The forecasts of the peak discharges were therefore represented through simulation mode values, i.e., with no data assimilation/updating. These 'simulation mode' forecasts can therefore be thought of as representing longer lead time forecasts where the effect of updating has died out.

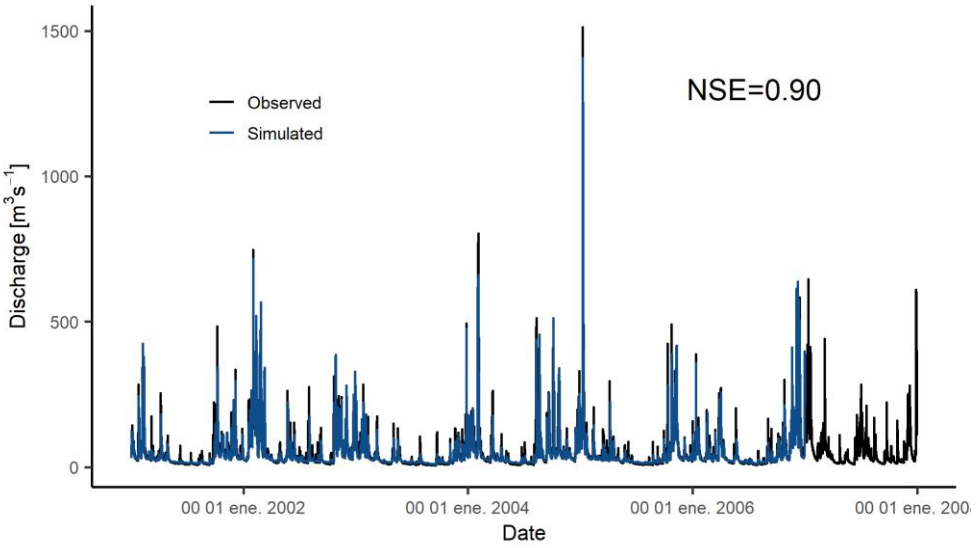


Figure 4.6: Simulated and observed discharges of the River Eden at Sheepmount. The simulated and observed discharge corresponds to the validation period (2001-2007). The NSE of the Eden SHETRAN model in the validation procedure was 0.90.

4.5.3 Selection of peak flows and their forecasts.

The peak flows y and their forecasts $\hat{y}$ are sampled in later chapters from a bivariate distribution with marginal distributions represented by a POT model. The variable y is here defined as the crest of the resulting hydrograph generated by a given rainfall event that crosses a given threshold level (y_b) of the river. The value of y_b must be set neither so high that only a few floods are considered in the hydrologic frequency analysis (HFA), i.e., it should be lower than the flooding threshold (y_T), nor so low that too many peaks are considered, which are not relevant to the HFA (see Figure 4.7).

As one can imply, the peak flow selection is strongly linked to y_b. The problem with y_b arises from the fact that the basic model hypothesis of any statistical model to be applied to y, the independence of consecutive peaks, can be affected when considering low thresholds (see, for example, the first peak in Figure 4.7). This

issue can be tackled by imposing further requirements on the basic criterion of choosing all the peak flows above y_b. These requirements are usually based on the area of the catchment, the time period between two consecutive peaks (t_P) and the difference between the magnitudes of the latter. For example, Cunnane (1979) imposes that consecutive peaks must be separated by three times the average time to peak of the hydrographs, or that the smallest discharge value between the magnitude of two consecutive peaks must be higher than two-thirds of the magnitude of the first peak. The Water Resources Council (USWRC, 1982) also suggests that two consecutive peaks are considered independent if they are separated by at least as many days as five plus the natural logarithm of square miles of the basin area. Other authors simply consider that consecutive independent peaks are characterized by a value of $t_P \geq 8$ or 15 days (Bogner et al, 2012; Karim et al, 2017).

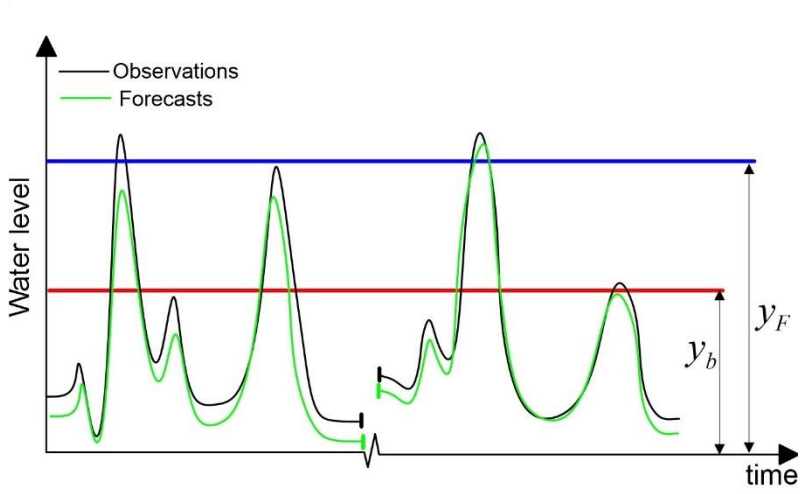


Figure 4.7: Illustration of y and $\hat{y}$ and the main thresholds (y_b and y_F) involved in the selection criterion for performing an HFA of these variates

On the other hand, $\hat{y}$ is defined as the crest of the forecast hydrograph used to forecast the peak flow y. The selection of $\hat{y}$ is, therefore, linked to the selected values of y where the forecast values might or might not fall above y_b. This could, for example, occur when the value of y is relatively close to y_b, such as the last peak in Figure 4.7. Thus, one has two criteria to select y and $\hat{y}$. The first one has to do with selecting the values of y and then selecting the values of $\hat{y}$ associate with y, and the other with selecting *only* the values of y whose associated values of $\hat{y}$ fall above y_b, such as the first three peaks shown in Figure 4.7. The latter option

increases the chances of using the same POT model (probability distribution) to represent y and $\hat{y}$.

Thus, the value of y_b considered in this work was a threshold level that has been set for all gauging stations used for flood warning purposes in England. When the water level of the river reaches that threshold, minor flooding is possible in the floodplain. Table 4.3 and Table 4.4 describes the value of this threshold for each gauging station analysed, whereas Figure 4.8 shows an example of the threshold value against actual readings for one of the gauging stations.

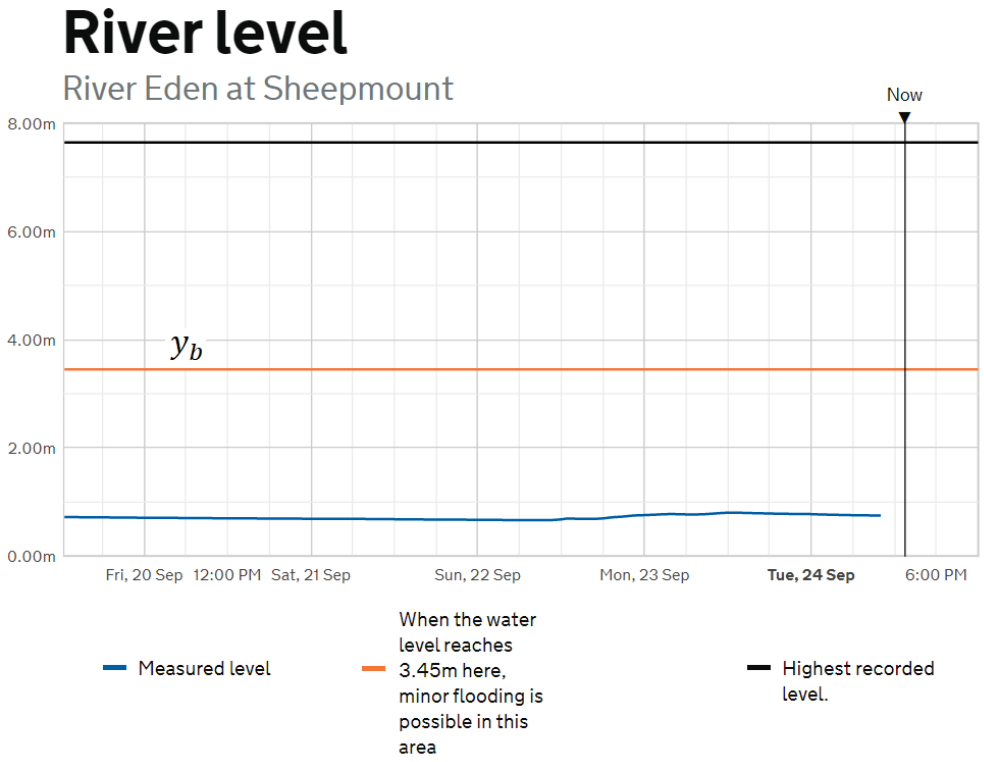


Figure 4.8: Value of y_b considered for one of the gauging stations considered in the analysis (River Eden at Sheepmount)
Source: (Environment Agency, 2019)

Table 4.3 Value of the threshold y_b and the resulting values of $\bar{y}$, s_y, and λ, for the gauging stations where only the values of y were available.

Code	Name	y_b [m]	$\bar{y}$ [m]	s_y [m]	λ
24003	River Wear at Stanhope	2.10	2.48	0.40	1.45
23006	River South Tyne at Featherstone	1.90	2.06	0.20	1.29
22007	River Wansbeck at Mitford	2.00	2.49	0.46	1.60

Table 4.4 Value of the threshold y_b for the gauging station River Eden at Sheepmount and the resulting values of $\bar{y}$, s_y,$\bar{\hat{y}}$, $s_{\hat{y}}$,and λ.

Code	Name	y_b [m]	$\bar{y}$ [m]	s_y [m]	$\bar{\hat{y}}$ [m]	$s_{\hat{y}}$ [m]	λ
76007	River Eden at Sheepmount	3.45	4.16	0.73	4.24	0.64	2.86

Once the values of y_b were defined for each gauging station, the values of y were selected, first defining the individual peaks above y_b, and then selecting the peaks, which were separated by at least eight days. For the case where the values of $\hat{y}$ were represented (River Eden at Sheepmount), the criterion for choosing the pairs $(y, \hat{y})$ was to select *only* the values of y whose associated values of $\hat{y}$ fall above y_b. Table 4.3 and 4.4 show the resulting value of the average number of peaks per year (λ), the means ($\bar{y}$ and $\bar{\hat{y}}$), and standard deviations (s_y and $s_{\hat{y}}$) of y and $\hat{y}$ when applying that criteria to each gauging station analysed, whereas Figure 4.9 illustrates some values of y for the river Eden at Sheepmount.

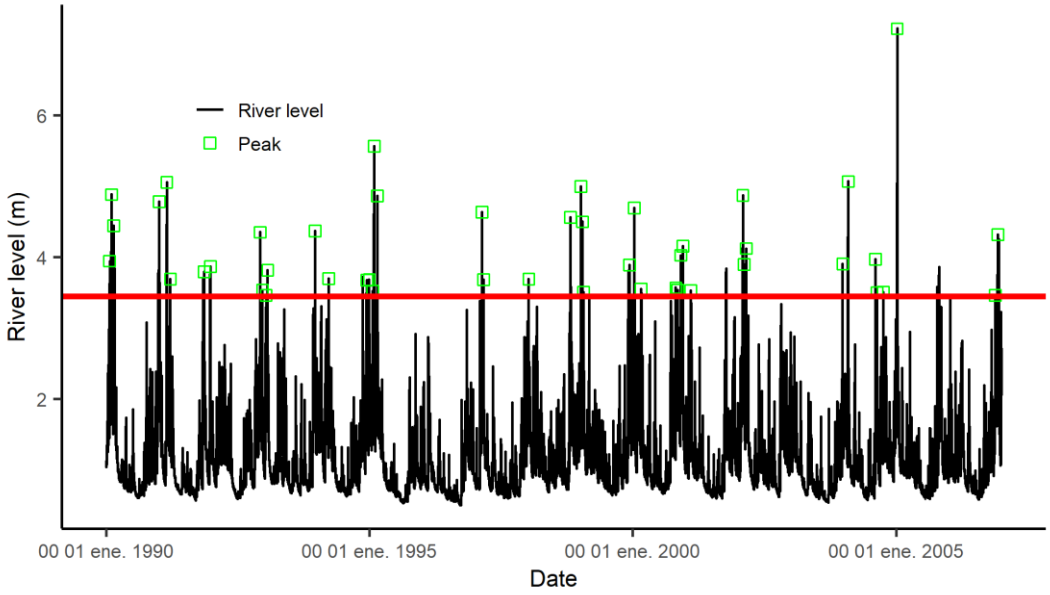


Figure 4.9: Hydrologic time series, and values of y for the period 1990:2006 obtained at the gauging station 76007 River Eden at Sheepmount
The green squares indicate the values of y and the red line the value of the threshold y_b.

The average number of peaks per year λ can be used to estimate the value of y associated with the return period of a Standard of protection (T_{SoP}). To compute y_{SoP}, first the probability of exceedance associated with T_{SoP} is estimated by:

$$P_{T_{SoP}} = \frac{1}{(T_{SoP})^{\lambda}} \qquad \text{Eq. 4.54}$$

Then, y_{SoP} is estimated as the $P_{T_{SoP}}$ quantile in the marginal distribution of y. This approach will be used to estimate important variables described in Chapter 5 and 6, such as the flooding threshold y_T.

4.5.4 Univariate analysis of y and $\hat{y}$

The aim of the univariate analysis of y and $\hat{y}$ was to analyse what the best probabilities distributions are to represent the marginal distributions in the bivariate analysis. It was done based on the GoF of the values of y and $\hat{y}$, for which, as was mentioned in section 4.2, three distributions were considered: the two-parameter exponential distribution, the three-parameter log-normal distribution, and the three-parameter Gamma distribution (see Table 4.1).

4.5.4.1 Parameter estimation

The distribution parameters can be estimated through different methods, such as the method of moments (MOM), the method of L-moments (ML), and the maximum likelihood estimation (MLE). In this work, the MOM was used, whose equations were previously described in section 4.2 (achieving high statistical efficiency in the estimates is not important here). To apply these equations, one can assume that the threshold parameter is known or unknown. The parameter estimation of the Lognormal distribution for the latter option is not straightforward (see section 4.2.2); thus, for the sake of simplicity, one can take the first option whose value can be assumed as the minimum value of the sample of y and $\hat{y}$. It is a common practice in a peak-over threshold analysis (Claps and Laio, 2003).

Table 4.5 and Table 4.6 shows the parameter estimates obtained when applying the moment equations described in section 4.2 and the above-mentioned assumptions.

Table 4.5 Estimation of the parameters of the three probabilistic models considered as candidates to represent y.

Code	Name	Location parameter	Exponential	Log-Normal		Gamma	
		y_o	$\widehat{\beta}_x$	$\hat{\mu}_z$	$\widehat{\sigma}_z$	$\widehat{\beta}_z$	$\widehat{k}_z$
24003	River Wear at Stanhope	2.12	0.36	-1.60	1.29	0.40	0.94
23006	River South Tyne at Featherstone	1.91	0.15	-2.33	0.97	0.24	0.64
22007	River Wansbeck at Mitford	2.03	0.46	-1.06	0.80	0.44	1.10
76007	River Eden at Sheepmount	3.46	0.70	-0.68	0.83	0.72	1.0

Table 4.6 Estimation of the parameters of the three probabilistic models considered as candidates to represent $\hat{y}$.

Code	Name	Location parameter	Exponential	Log-Normal		Gamma	
		$\widehat{y}_o$	$\widehat{\beta}_{\hat{x}}$	$\hat{\mu}_{\hat{z}}$	$\widehat{\sigma}_{\hat{z}}$	$\widehat{\beta}_{\hat{z}}$	$\widehat{k}_{\hat{z}}$
76007	River Eden at Sheepmount	3.47	0.77	-0.50	0.70	0.50	1.56

4.5.4.2 Goodness of fit (GoF) of y

A Goodness of Fit test (GoF) can be based on a visual inspection and/or through statistical GoF tests (D'Agostino and Stephens, 1986; Kottegoda and Rosso, 2008). A common approach for the first one has to do with plotting the theoretical CDF against the empirical one. For this purpose, the empirical probabilities can be computed by using the Weibull equation.

$$F_m = \frac{m}{n+1} \qquad \text{Eq. 4.55}$$

where F_m is the probability of non-exceedance of the event m, which is defined through the rank of descending values, and n is the sample size.

Since the visual inspection of GoF can be quite subjective, the computation of statistical GoF tests is usually suggested, which can be used to know whether or not it is reasonable to assume that the random sample of y comes from an assumed

probability distribution based on a hypothesis testing where the null and alternative hypotheses are:

H_0: Sample data come from the assumed probability distribution.

H_A: Sample data do not come from the assumed probability distribution.

The null hypothesis is rejected if the probability of the computed value, i.e., the p-value, is lower than a defined level of significance. The level of significance considered in this work was 5%.

When the population parameters are unknown and must be estimated by sample statistics, the Lilliefors-corrected Kolmogorov-Smirnov (correct-KS) test can be used to check the model assumption (Lilliefors, 1969; Crutcher, 1975). These values were estimated by using the R package "KScorrect" (Novack-Gottshall and Wang, 2019), which is based on MLE.

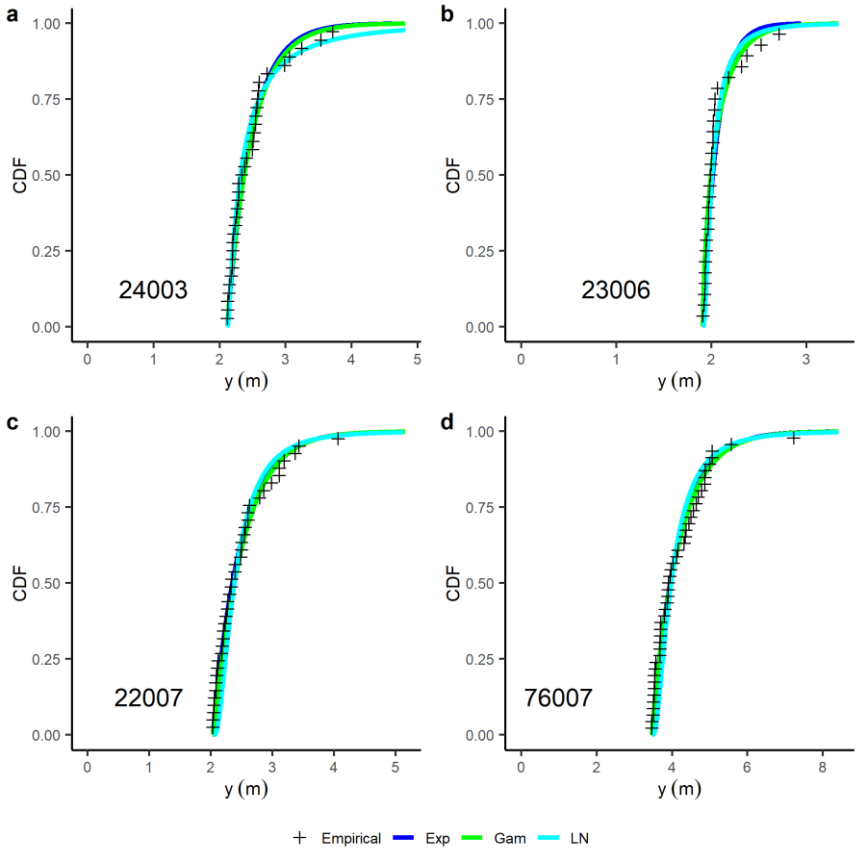


Figure 4.10: Comparison of the empirical and theoretical CDF for the values of *y* obtained for each gauging station

The GoF results of y are shown in Table 4.7 and Figure 4.10. As one can see, only for one of the gauging stations (23006 River South Tyne at Featherstone), the statistical GoF test indicates the suitability of one of the probabilistic models (Log-normal distribution) by rejecting the null hypothesis of the others. For some samples, one must, therefore, choose one probability distribution among the ones that pass the statistical test. When this occurs, it is usually suggested to consider the complexity of the probabilistic model as a discriminatory factor. Therefore, for these cases, the Exponential distribution can be considered as the main option to represent the values of y given the low complexity of the univariate probabilistic model (two parameters: see Table 4.1). However, any of the distributions accepted under the null hypothesis for each station can be used.

Table 4.7 Statistical GoF of y based on the correct-KS test for each gauging station
In this table *Yes* means that the null hypothesis was not rejected, i.e., the sample data passed the test. *No* means the null hypothesis was rejected, i.e., the sample data did not pass the test.

Code	Name	Kolmogorov Smirnov test		
		Lognormal	**Gamma**	**Exponential**
24003	River Wear at Stanhope	*Yes*	*Yes*	*Yes*
23006	River South Tyne at Featherstone	*Yes*	*No*	*No*
22007	River Wansbeck at Mitford	*Yes*	*Yes*	*Yes*
76007	River Eden at Sheepmount	*No*	*Yes*	*Yes*

4.5.4.3 Goodness of fit (GoF) of $\hat{y}$

As was mentioned before, the GoF of $\hat{y}$ was only done for the gauging station 76007 River Eden at Sheepmount. The results based on the correct-KS test and the comparison of the empirical and theoretical CDF are shown in Table 4.8 and Figure 4.11, respectively. As one can see, the null hypothesis is not rejected for all the assumed probability distributions, so the same conclusion drawn above for the case of y applies here.

Table 4.8 Statistical GoF based on the correct-KS test for the values of $\hat{y}$ corresponding to the values of y obtained from the gauging station 76007 River Eden at Sheepmount.
In this table *Yes* means that the null hypothesis was not rejected, i.e. the sample data passed the test. *No* means the null hypothesis was rejected, i.e. the sample data did not pass the test.

Variable	Kolmogorov Smirnov test		
	Lognormal	**Gamma**	**Exponential**
Values of $\hat{y}$ corresponding to the values of y obtained from the gauging station 76007 River Eden at Sheepmount.	*Yes*	*Yes*	*Yes*

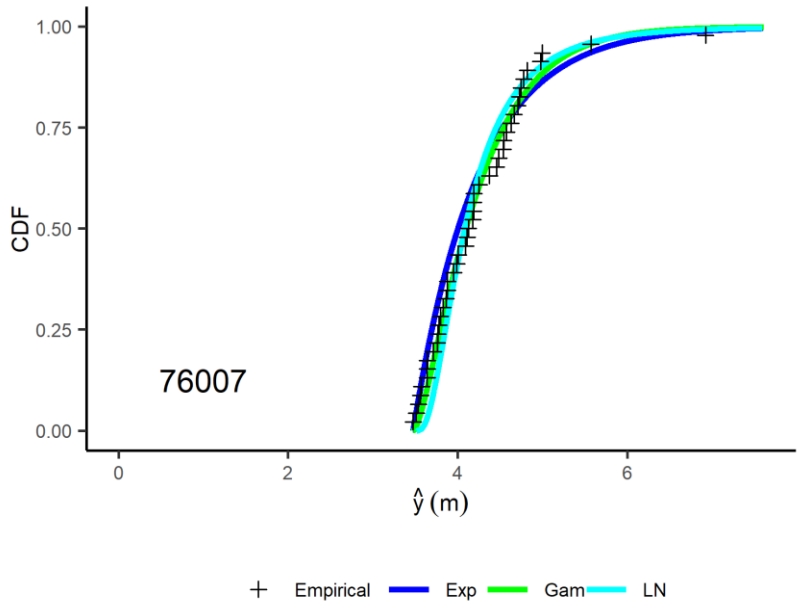


Figure 4.11: Comparison of the empirical and theoretical CDF for the values of $\hat{y}$ corresponding to the values of y obtained from the gauging station 76007 River Eden at Sheepmount

4.5.5 Bivariate analysis of y and $\hat{y}$ based on the Moran-Downton BED

In terms of GoF, the principle of bivariate frequency analysis should be the same as in the case of single-variable frequency analysis. Although the statistical GoF test for bivariate frequency analysis is not straightforward, an intuitive investigation of the agreement between empirical and theoretical joint probabilities such as the approaches used in the work of Yue (2000) and Yue (2001) might be useful for checking the suitability of any bivariate probabilistic model. In this approach, n bivariate observations (s_i, q_i), are arranged in ascending order, thus obtaining pairs (s_j, q_j). The empirical non-exceedance joint probabilities are then computed as:

$$\hat{F}(s_j, q_j) = \frac{\#pairs\ (s_i, q_i)\ with\ s_i \leq s_j\ and\ q_i \leq q_j}{n+1} \qquad \text{Eq. 4.56}$$

$\hat{F}(s_j, q_j)$ can be then compared with the theoretical joint probabilities of $F(S \leq s_j, Q \leq q_j)$ of the bivariate model to be tested.

In this work, that method was applied to the sample of the pairs $(y,\hat{y})$ obtained from the gauging station River Eden at Sheepmount. The aim was to test the assumption that the sample of pairs $(y,\hat{y}) \sim BED(\beta_x, \beta_x, y_o, \hat{y}_o, \rho_{yy})$ is consistent with the

Moran-Downton BED. Figure 4.12 shows the estimation of $\rho_{y\hat{y}}$ based on the sample correlation coefficient between y and $\hat{y}$, whereas Table 4.9 shows the estimated MOM values of the parameters of the five-parameter BED. Figure 4.12 shows that the values of $\hat{y}$ show a relatively small error for high values of y which is not the bahaviour commonly observed when making forecasts. However, the upper range of values is poorly sampled because of the small sample size, so this can have happened by chance, but also perhaps through overfitting to the highest peaks when calibrating the SHETRAN model.

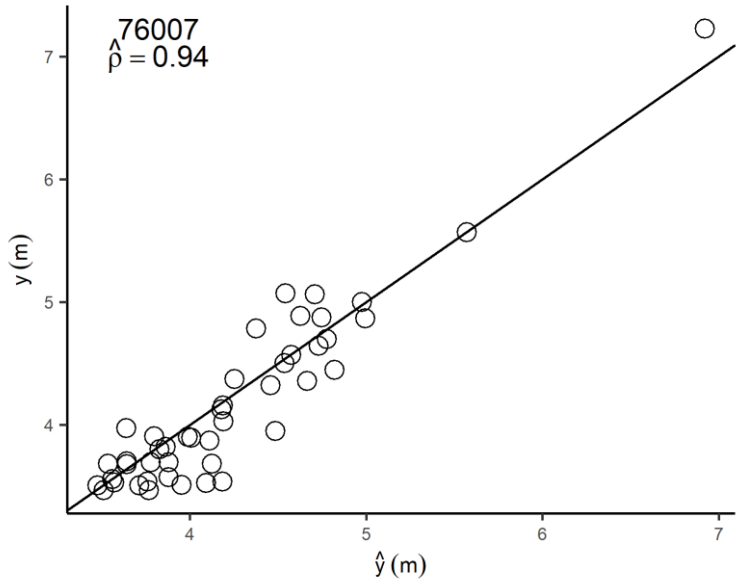


Figure 4.12: plot $\hat{y}$ versus y for the values obtained from the gauging station 76007 River Eden at Sheepmount

Table 4.9 Estimated values of the parameters of the five-parameters BED for the gauging station River Eden at Sheepmount.

Code	Name	y_o	$\hat{y}_o$	$\hat{\beta}_{\hat{x}}$	$\hat{\beta}_x$	$\hat{\rho}_{y\hat{y}}$
76007	River Eden at Sheepmount	3.46	3.47	0.77	0.70	0.94

The bivariate GoF of the pairs (y,y) based on the Moran-Downton BED was thus based on the following steps:

- ✓ **Step 1**: Based on the estimated values of the parameters, compute the standardized exponential variables of the pairs (y,y) as:
  $v = \frac{y-y_o}{\hat{\beta}_x}$ *and* $\hat{v} = \frac{\hat{y}-\hat{y}_o}{\hat{\beta}_{\hat{x}}}$
- ✓ **Step 2:** Arrange the observations $(v_i, \hat{v}_i)$ in ascending order and obtain the pairs $(v_j, \hat{v}_j)$.
- ✓ **Step 3**: Compute the empirical non-exceedance joint probabilities of the pairs (v_j,$\hat{v}_j$), i.e. $\hat{F}(v_j,\hat{v}_j)$, based on Eq. 4.56.

- ✓ **Step 4**: By using the relationship $\rho_{y\hat{y}} = \rho_{v\hat{v}}$, compute the theoretical joint probabilities of the pairs $(v_j,\hat{v}_j)$, i.e., $F(v_j,\hat{v}_j)$, based on Eq. 4.34.
- ✓ **Step 5:** Do a bivariate GOF via visual inspection by plotting in the x-axis the order of the pairs $(v_j,\hat{v}_j)$ and in the y-axis, the values of $\hat{F}(v,\hat{v})$ and $F(v,\hat{v})$.

The results of the above-mentioned steps are shown in Figure 4.13. As one can see, the plot indicates that there is no significant difference between the observed and theoretical probabilities. Therefore, one can conclude that the five-parameters BED is suitable to represent the joint probabilities of the pairs $(y,\hat{y})$ obtained from the gauging station River Eden at Sheepmount.

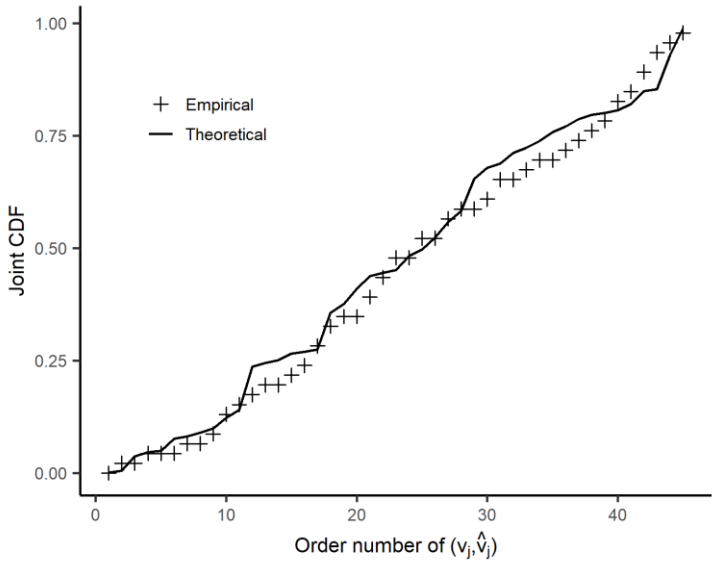


Figure 4.13: Bivariate GoF of the pairs $(y,\hat{y})$ based on the Moran-Downton BED for the values obtained from the gauging station 76007 River Eden at Sheepmount

Based on the outcome of this test, it is assumed that observed peak levels, and their forecasts from a calibrated model with a relatively high NSE and correlation, can be described by the BED model. Moreover, it is reasonably assumed, although not proven here, that any of the candidate distributions that passed the GOF test for y at the other sites can be used to describe $\hat{y}$, thereby allowing corresponding bivariate distributions to be used for simulation experiments in later chapters.

4.6 Forecast performance

The prior sections showed how to represent pairs ($\hat{y}$,y) with a given dependence structure and associated errors through different bivariate probabilistic models; it provides the basis of the MCFG. This also provides the starting point for the sensitivity analysis used to explore the importance of the different factors influencing the performance of a FEWS. However, the description of the dependence structure of pairs ($\hat{y}$,y) through a correlation coefficient is not enough since one needs to associate it with a given forecast horizon or lead time (τ), and a performance measure, which is a common characteristic used to define a forecasting model. A MCFG can be set fully if these two characteristics are taken into account. This section, therefore, explains how the dependence structure of pairs ($\hat{y}$,y) was associated with τ and the forecasting performance.

It is well known that there is a trade-off between the forecast performance and the forecast horizon or lead time τ (Blöschl, 2008; Parker, 2017).The correlation $\rho_{y\hat{y}}$ is therefore necessarily associated with a given τ. In this research, that relationship was represented through a subjective lead-time-correlation function (see Figure 4.14) analogous to that found by Schröter *et al.* (2008). This function describes the common behaviour of forecasting models forced with precipitation. That is, it represents the fact that, for τ values lower than basin lag time (L), with a well-calibrated hydrological model, the performance of a forecasting model is relatively high, as the forecasts are based on observed precipitation by using, for example, gauge-based quantitative precipitation estimation (QPE). Moreover, forecast updating will improve the correlation for short lead times. Past this value L, the forecasting model has to be forced with quantitative precipitation forecasts (QPFs), and its performance drops monotonically. The slope of this function before L defines the quality of forecasting models based on QPE, such us models based on gauge-based QPE or gauge-radar based QPE, and the slope after L defines the quality of the forecasts based on QPFs. This function is defined by:

$$\rho_{y\hat{y}} = \begin{cases} -0.015 * (\tau - L) + IP_\rho & \tau \leq L \\ -0.030 * (\tau - L) + IP_\rho & \tau > L \end{cases} \qquad \text{Eq. 4.57}$$

where IP_ρ is the value of $\rho_{y\hat{y}}$ at the inflection point of the correlation-lead-time function and represent the correlation associated to a catchment lag L. In Eq. 4.57,

L and τ are defined in the same units, e.g., hours. The sensitivity of forecast performance to perturbations in the IP_ρ value is explored in Chapters 5 and 6.

Biases in the mean and variance of forecasts relative to the observed values also affect the forecast performance, and these are explored in the sensitivity experiments in Chapter 5.

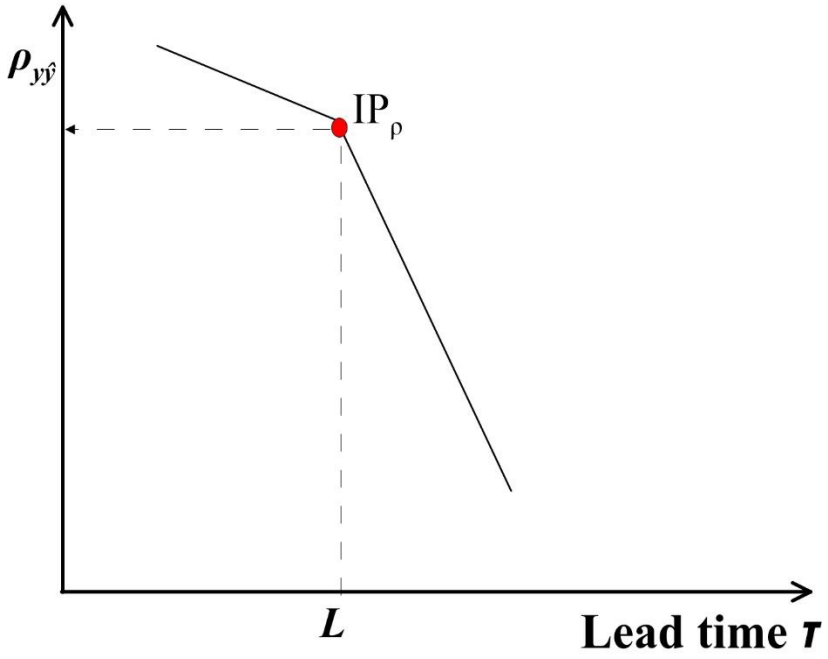


Figure 4.14: Lead-time-correlation function used to represent the forecast performance as a function of lead time.

4.7 Sensitivity analysis

In Chapters 5 and 6, sensitivity analyses (SA) of the factors controlling the performance of a FEWS are performed. The basic approach to the SA is to state a baseline scenario and then test several scenarios which take into account the main assumptions in the estimation of the performance. The forecast performance was one of them, which, as was explained above, is linked with the setting of the MCFG parameters. Therefore, this section of the chapter describes the parameters used to represent the MCFG and which of them were considered in the SA.

The baseline scenario considered the results obtained for the gauging station River Wansbeck at Mitford as reference values. Based on the results in Table 4.7, it is assumed that y and $\hat{y}$ can be represented by the same Lognormal POT model type, which, in turn, leads to the assumption that the pairs $(y,\hat{y}) \sim BLND(\mu_z, \sigma_z, \mu_{\hat{z}}, \sigma_{\hat{z}}, y_o, \hat{y}_o, \rho_{z\hat{z}})$. Furthermore, the baseline model assumes

an ideal case of the forecasting system where the marginal distribution of the observations is equal to that of their forecasts; therefore, the moments of $\hat{y}$ are equal to those of y. The lead time τ was assumed to be equal to the catchment lagtime L, which, in turn, was assumed to be equal to 6 hours. Finally, the IP_ρ value was assumed to be 0.85; the justification for this is discussed in Chapter 5. Based on this analysis, the MCFG was represented by eleven parameters in the SA framework, where five of them represent the characteristics of the basin and the others the characteristics of the forecasting model. Table 4.10 shows these parameters and their assumed values for the baseline scenario.

Table 4.10: Assumed input parameter values for the MCFG in the baseline generic fluvial case

Abbreviation	Description	Value adopted	Parameter associated with:
μ_y	Mean of y.	2.51	The river basin
σ_y^2	The variance of y.	0.20	
y_o	Location parameter of y	2.03	
L	Basin lagtime L	6 hrs	
$\mu_{\hat{y}}$	Mean of the forecasts of y	μ_y	Flood forecasting system
$\sigma_{\hat{y}}^2$	The variance of the forecasts of y	σ_y^2	
$\hat{y}_o$	Location parameter of the forecasts of y	y_o	
IP_ρ	The inflection point of the lead time-performance function	0.85	
τ	Lead time	6 hr	

4.8 Main findings

The architecture of the MCFG is novel; it allows forecast uncertainty to be linked with a forecasting lead time in the simulated generic fluvial case. The univariate analysis of POT data for y and $\hat{y}$ at selected gauging stations suggested that the probability distribution type of $\hat{y}$ is that of y, and that a number of different distributions could be used to describe y. This allowed the assumption to be made that the five-parameter bivariate exponential distribution (BED), the seven-parameters bivariate log-normal distribution (BLND), and the seven-parameter bivariate gamma distribution (BGM) are suitable models to represent the pairs $(y, \hat{y})$. Algorithms for sampling bivariate pairs of y and $\hat{y}$ from two of these bivariate distributions (BED and BLND) were presented. The algorithm relates to BGM will be described in Chapter 7.

A formal criterion for selecting the threshold y_b defining the peak water levels in the POT frequency analysis was proposed.

Chapter 5. Evaluation of flood warning performance in terms of reliability

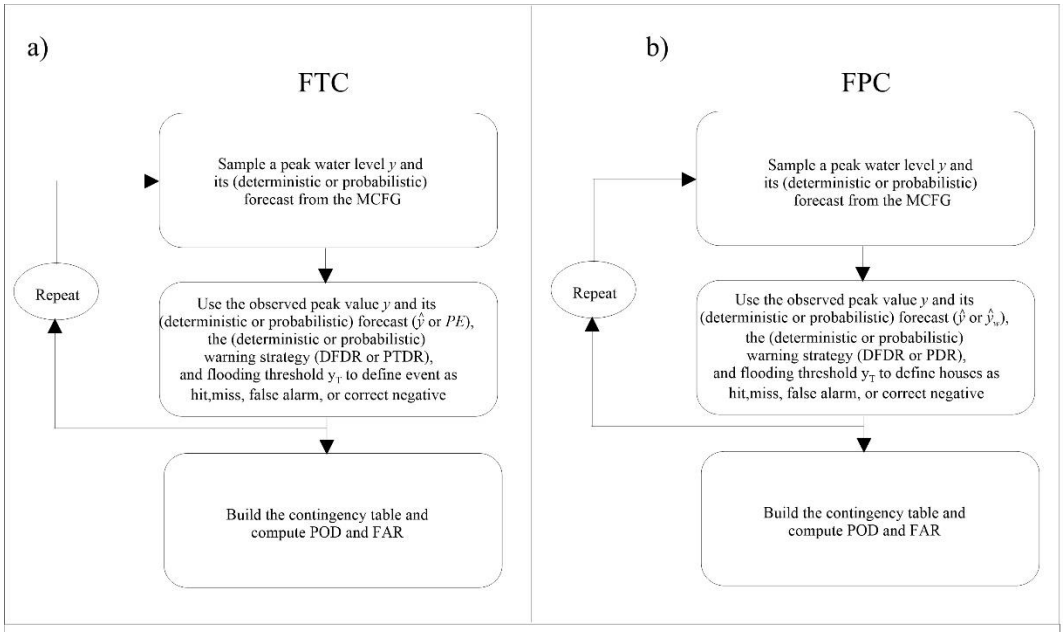
There are a considerable number of acronyms in this Chapter. Therefore, the Chapter starts with a list of them with the aim of helping the reader from the beginning to be familiar with the concepts used in the analysis of this part of the thesis. The concepts of several acronyms have already been introduced in the prior Chapters; therefore, the list of acronyms can also be used by the reader to navigate the thesis to look for more detailed information.

Acronym	Meaning
BLND	Bivariate Log-normal distribution
DFDR	Deterministic forecast-based decision rule. Decision rule used in a deterministic FEWS to represent warning decisions based on the values of $\hat{y}$ (section 3.2.2). Used in conjunction with the Floodplain property-based criterion (FPC) and the Floodplain threshold-based criterion (FTC).
FAR	False alarm ratio. Skill score used to define the reliability of flood warnings defined as the ratio of the number of hits and the sum of hits and false alarms (section 2.7.1).
FEWS	Flood early warning system
FPC	Floodplain property-based criterion. Criterion to evaluate the reliability of flood warnings in terms of warned and flooded houses. Values of FAR and POD can be derived from this criterion (section 2.7.1).
FWDC	Flood warning decision component. Component of the generic framework (section 3.2.1)
FTC	Flooding-threshold criterion. Criterion to evaluate the reliability of flood warnings based on whether the occurrence or non-occurrence of a flood event in the at-risk community was preceded by a warning. Values of FAR and POD can be derived from this criterion (section 2.7.1).
NSE	Nash-Sutcliffe Efficiency
MCFG	Monte Carlo flood and forecast generator. Component of the generic framework (section 3.2.1)
PE	Probability of exceedance of a predefined level derived from the conditional distribution of y given $\hat{y}$, i.e., $f(y\|\hat{y})$. Variate used by the probabilistic rules PTDR and PDR (section 3.2.2)
PDR	Probabilistic decision rule. Decision rule used in a probabilistic FEWS to represent warning decision based on a warning level $\hat{y}_w$ derived from $f(y\|\hat{y})$ (section 3.2.2) Used in conjunction with the Floodplain property-based criterion (FPC).
POD	Probability of detection. Skill score used to define the reliability of flood warnings defined as the ratio of the number of hits and the sum of hits and misses (section 2.7.1).
PT	Probabilistic threshold used for the probabilistic rules PDR and PTDR (section 3.2.2).
PTDR	Probabilistic-threshold-based decision rule. Decision rule used in a probabilistic FEWS to represent warning decision based on the PE of the flooding threshold y_T derived from $f(y\|\hat{y})$ (section 3.2.2). Used in conjunction with the Floodplain threshold-based criterion (FTC)
PU	Predictive uncertainty.
RIC	Response and impact component. Component of the generic framework (section 3.2.1)
RRA	Risk-reduction action. A proactive action conducted after a warning has been issued.
SoP	Standard of protection.

5.1 Introduction

In the Monte Carlo (MC) framework used to explore the performance of a flood early warning system (FEWS) for the generic fluvial case (Chapter 3), the peak river levels (y) and their forecasts ($\hat{y}$) are generated from the MC flood and forecast generator (MCFG). Chapter 4 gives a detailed description of how this component of the framework can create pairs $(y, \hat{y})$ and the predictive uncertainty (PU) associated with $\hat{y}$. Different bivariate probabilistic models are assumed, and the dependence structure of this model (represented by the correlation coefficient $\rho_{y\hat{y}}$) is associated with a lead time (τ) through a lead time-correlation function. In this component, a return period is also associated with each value of y.

In this chapter, a full description of the second and part of the third component of this MC framework is provided. The former component is called the flood warning decision component (FWDC) and represents the decisions made by a Warner (a word adopted in this research to refer to a person involved in preparing and issuing flood warnings, usually a member of a governmental entity) based on the forecasts generated by the MCFG and a flooding threshold (y_T) associated with an at-risk community. The third component, i.e., the response and impact component (RIC), is used partially to only represent the forecast and observed impact in terms of affected houses. The adopted bivariate model used in the MCFG to represent the peak river levels y and their forecasts $\hat{y}$ is the bivariate Lognormal distribution (BLND). The reliability or credibility of a FEWS is the flood warning performance characteristic to be explored in this chapter. This analysis depends on warning decisions and the uncertainty of the forecasts and requires the addition of two parameters to the ones considered in the MCFG. These parameters are associated with the characterization of the at-risk community and have to do with the return period of its standard of protection (SoP) (T_{SoP}), and the number of houses at risk in the floodplain ($n_{houses}^{at-risk}$). How these parameters are included in the MC framework is explained later. The generic framework (Figure 3.1) is, thus, partially used to build a MC framework to explore the reliability of flood warnings (Figure 5.1). Thus, in this case, the response is left out, which is the basis of its economic component that, in turn, allows the performance of a FEWS to be explored in terms of the *economic effectiveness*. This analysis will be done in Chapter 6.



FTC: Flooding-threshold criterion. Criterion to evaluate the reliability of flood warnings based on whether the occurrence or non-occurrence of a flood event in the at-risk community was preceded by a warning.
FPC: Floodplain property-based criterion. Criterion to evaluate the reliability of flood warnings in terms of warned and flooded houses.
MCFG: Component of the MC framework to generate pairs $(y,\hat{y})$ with its associated conditional distribution $f(y \mid \hat{y})$.
$\hat{\mathbf{y}}$: Decision variable in a determinsitic scenario. Used in conjunction with FTC and FPC.
PE: Decision variable in a probabilistic scenario. Probability of exceedance of the flooding threshold y_T derived from the conditional distribution of y given $\hat{y}$, i.e., $f(y \mid \hat{y})$. Used in conjunction with FTC.
$\hat{\mathbf{y}}_{\mathbf{w}}$**:** Decision variable in a probabilistic scenario. Forecast level derived from $f(y \mid \hat{y})$ as the probability of exceedence of probabilistic threshold PT.
DFDR: Deterministic forecast-based decision rule. . Decision rule used in a deterministic scenario to represent warning decision based on the values of $\hat{y}$. Used in conjunction with FPC and FTC.
PTDR: Probabilistic-threshold-based decision rule. Decision rule used in a probabilistic scearion to represent warning decision based on the PE. Used in conjunction with FTC.
PDR: Probabilistic decision rule. Decision rule used in a probabilistic scenario to represent warning decision based on a warning level $\hat{y}_w$ derived from $f(y \mid \hat{y})$. Used in conjunction with FPC.

Figure 5.1: Illustration of the MC framework to explore the reliability of flood warnings. The MC framework explores the flood warning reliability estimated based on the flooding threshold-based criterion FTC (a) and floodplain property-based criterion FPC (b) by using a (deterministic or probabilistic) warning strategy.

The reliability of a FEWS, which, according to Parker (2017), is associated with the presence of false and missed alarms, is evaluated through the skill scores known as the probability of detection (POD) and false alarm ratio (FAR). This flood warning performance attribute is captured based on a contingency-table-based method that represents the outcomes of deterministic-based and probabilistic-based warning decisions. Contingency tables are built based on two criteria: a flooding threshold-based (FTC) and floodplain property-based (FPC) criterion. The former assesses the performance of a FEWS based on whether the occurrence or non-occurrence of a flood in the at-risk community was preceded by a warning, whereas the latter based on whether a warned property was or was not subsequently flooded (Figure 5.1). The FTC is a well-known criterion used in several research works (Verkade and Werner, 2011; Pappenberger *et al.*, 2015; Bischiniotis *et al.*, 2019), whereas the FPC is a contribution of this research which gives a more consistent estimation of flood warnings' reliability because it is estimated based on warned and flooded houses.

The MC framework is used to explore the *reliability* of a FEWS through several sensitivity experiments. To do so, a set of values for the input parameters is first assumed to characterise the baseline scenario, where the at-risk community, the river system, and the flood forecasting system are characterised. Then, the sensitivity of the *reliability* to perturbations of the main parameters is fully explored through a one-at-time method.

This chapter is structured as follows: In section 5.2, the rationality of the FTC and FPC used to define the performance of a FEWS is explained based on the warning criterion used in two types of FEWSs. Then, in section 5.3, the simulation of the target information for evaluating the performance of a FEWS based on the FPC is described. Next, in section 5.4, the contingency-table-based method for the FTC and FPC used to evaluate the reliability of flood warnings is explained. After this, in section 5.5, the skill scores used to describe the performance of a FEWS are introduced, and, in section 5.6, the decision rules used to describe the FWDC based on different types of forecast information are explained. Finally, in section 5.7, the results of the sensitivity experiments are shown.

5.2 Flood warning criteria for fluvial floods

The FTC and FPC utilised in this work to explore the performance of a FEWS have been defined based on the warning criterion usually used in two types of FEWSs. Based on the flood forecasting technique they use, they have been called in this work a hydrologic forecast-based FEWS and an inundation forecasting-based FEWS. Table 5.1 gives a brief description of these systems, and Figure 5.2 illustrates the warning criterion for each of them. They are explained as follows.

The flood forecasting techniques applied in hydrologic forecast-based FEWS are usually based on hydrological and/or flood routing models (hydrologic or hydraulic models) and offer a relatively simple approach for flood predictions. They have been the dominant choice of river flood warning systems, where a threshold-based approach for issuing a warning is normally used. The FTC evaluates the performance of this type of FEWS where the warning criterion is based on a flooding threshold approach. In this case, forecast water levels at a river section in or close to the at-risk community are compared with a pre-defined flooding threshold y_T, which is often set based on experience, a pre-defined return period,

historical data, and/or detailed hydraulic modelling of river response. Here, y_T is assigned a return period. The warning decision is based on the direct comparison between the forecasts and y_T. Figure 5.2a shows an example when considering a deterministic forecast; in this case, a warning is issued when the forecast indicates the exceedance of this threshold. This type of FEWS does not target warnings on individual properties or small areas within the at-risk community but aims to alert the flood authority of a potential flood that, in turn, uses a risk-reduction action (RRA) to mitigate the flood impact. A common RRA in this type of FEWS is raising demountable defenses, which can need a 24-h warning lead-time to be erected. Flooding threshold-based warning decisions have been the basis of some important research (Verkade and Werner, 2011; Dale *et al.*, 2014).

Table 5.1 Description of two types of FEWS

Type of river flood warning system	Aim	Warning decision criterion	Risk reduction action (RRA)
Hydrologic forecast-based FEWS	Issue flood warnings to flood authority	Direct comparison between forecasts and the flooding threshold y_T	Raising demountable defenses
Inundation forecasting-based FEWS	Issue flood warnings to vulnerable properties in the flood risk community		Moving house contents upstairs and/or evacuating them and the residents

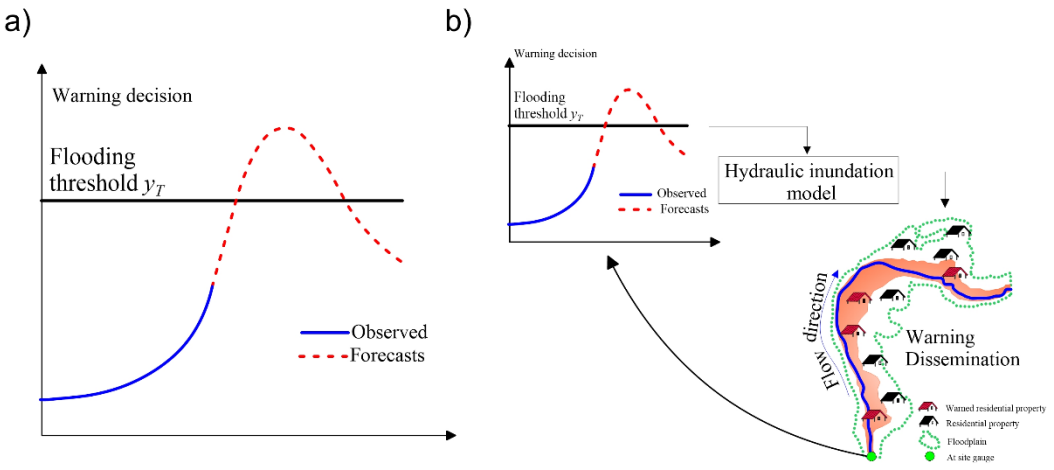


Figure 5.2: Illustration of the warning criterion for a flooding threshold-based (a) and inundation forecasting-based (b) warning system.

The inundation forecasting-based FEWS is considered to be the most sophisticated flood warning service (WMO, 2011). The flood forecasting system of this FEWS is usually made up of a hydrological and hydraulic inundation model, e.g., a

hydrodynamic model, where the output of the former is used to feed the latter to forecast the floodplain inundation. Even though these flood forecasting systems are considered complex and computationally expensive (particularly for a catchment-wide integrated hydrological and hydrodynamic modelling approach), with the development of high-performance computing techniques, it is now feasible to use these flood forecasting techniques in flood warning services (Ming *et al.*, 2020). The FPC evaluates the performance of this type of FEWS. In this case, the warning decisions can also be based on the direct comparison between forecast water levels and the flooding threshold y_T, and the forecast of the flood extent can be used to disseminate flood warnings (Figure 5.2b). Vulnerable properties are identified by intersecting the maps of the property locations and the forecast inundation area. In this context, an inundation forecasting-based FEWS aims to target warning on vulnerable properties in the at- risk community, and the RRA, in this case, is done by houseowners, which can be, for example, by mounting flood defences for individual flood properties and/or moving house contents upstairs or evacuating contents and residents.

5.3 Simulation of the flooded and warned properties

As was explained above, evaluating the performance of a FEWS based on the FPC is one of the aims in this chapter. To the best of the author's knowledge, there is not a work that has reported such an analysis. The target information for this task is the flooded and warned properties in the flood risk community after a flood event occurs. A rational method to do that would be one that represents: i) the infrastructure of the at-risk community, ii) the observed and forecast hydrograph of potential floods event at a river section in or close to the at-risk community, and iii) the associated observed and forecast flood extent through a hydraulic inundation model (Figure 5.3a). The MC framework represents this rational approach indirectly. The aim of this section of the chapter is, therefore, to explain this indirect method.

To quantify the number of flooded properties ($n_{houses}^{flooded}$), the MC framework uses an assumed *impact curve* for the flood risk community, i.e., a curve that relates the return period of a current or future flood event to the magnitude of the impact (Sayers *et al.*, 2018); in this case, the number of properties that would be flooded. This type of curve has been used for performing a national flood risk assessment in

the United Kingdom through the UK Future Flood Explorer (Sayers *et al.*, 2015); a flexible tool to simulate flood risk scenarios throughout the country.

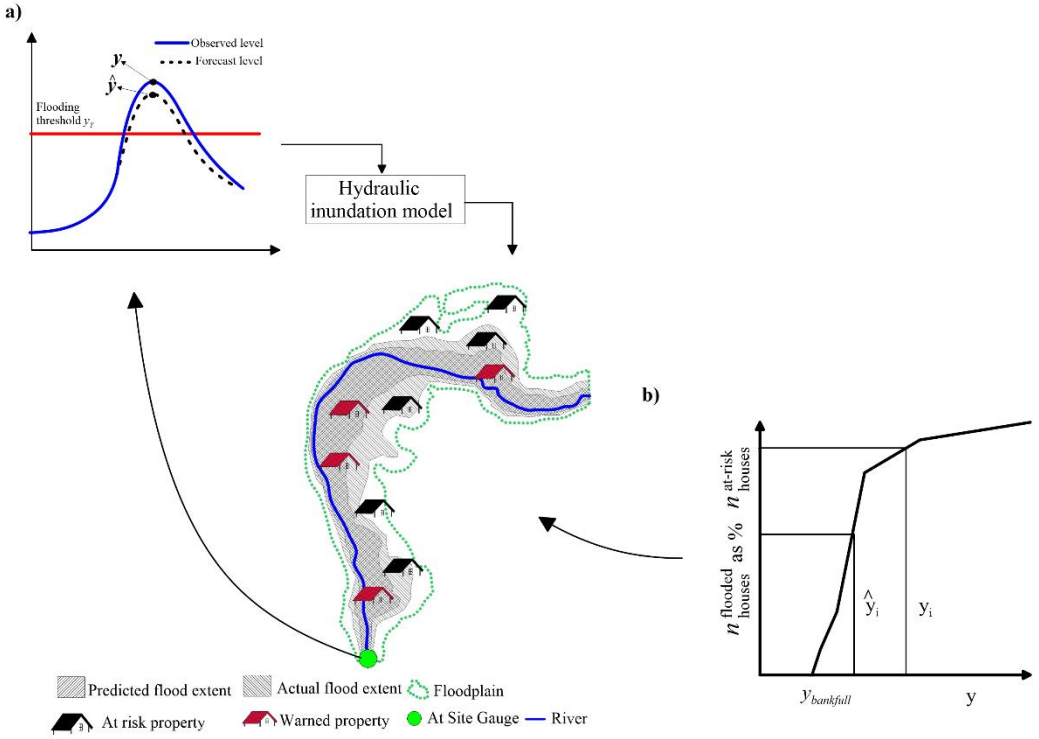


Figure 5.3: Illustration of the impact curve used to quantify the number of flooded and warned houses.

Figure 5.3b shows the impact curve used by the MC framework. The values of this impact curve were taken from Table 4.6 of the Multi-Coloured Manual Handbook (Penning-Rowsell *et al.*, 2020), which suggests using these values in a method to estimate the economic damage in a floodplain. This method is used when knowledge of flood depth and its associated return period is poor, but the Standard of Protection SoP, property type, and numbers are known. The number of houses flooded was set to zero at the water level that defines the bankfull condition; the return periods reported in the literature for the latter appear to fall somewhere in the range1-4 years based on annual maximum series (Williams, 1978; Andrews, 1980; Petit and Pauquet, 1997; Castro and Jackson, 2007; Ahilan *et al.*, 2013). These values would be longer if partial-duration series are considered as it contains more flood events than the annual maximum series(Edwards *et al.*, 2019). In this context, a 2.5 year-return period was assumed in this work to represent the bankfull condition.

The assumed impact curve (Figure 5.3b) gives an estimate of the number of properties affected by different floods by taking as reference the number of properties flooded by a 200-year flood ($n_{houses}^{200-year-flood}$). The MC framework assumes that the floodplain is defined by a 200-year flood extent. The total number of houses at risk $n_{houses}^{at-risk}$ is, therefore, assumed to be affected by a 200-year flood extent, i.e., $n_{houses}^{at-risk} = n_{houses}^{200-year-flood}$. Thus, $n_{houses}^{flooded}$ are computed based on the return period of y by taking as reference $n_{houses}^{at-risk}$ that is an input parameter of the MC framework.

The assumed impact curve is also used to estimate the number of warned properties (n_{houses}^{warned}), which represents the vulnerable properties warned by the inundation forecasting-based FEWS (Figure 5.2b). The number of houses warned n_{houses}^{warned} in the MC framework is estimated as a function of $\hat{y}$ by interpolating this value on the impact curve. With this approach, one tries to explore/represent the uncertainty of the inundation forecasting related to the magnitude of the flood event (peak water level). Figure 5.3b illustrates the computation of n_{houses}^{warned} and $n_{houses}^{flooded}$ for a pair $(y, \hat{y})$ based on the impact curve-based method and what they would represent in reality.

5.4 Analysis of the flood warning performance

Parker (2017) defines reliability as one of the common flood warning performance characteristics. This attribute of the FEWS is associated with the presence of false and missed alarms. The skill scores known as the probability of detection (POD) and the false alarm ratio (FAR) are often used as reliability measures which, in essence, give a "snap-shot" of this attribute of the FEWS (see section 2.7.1)

Based on this concept, a contingency table-based method was used to define the reliability of a FEWS based on the FTC and FPC. In the former, it is evaluated based on whether the occurrence or non-occurrence of a flood event in the flood risk community was preceded by a warning. In the latter, the reliability is, on the other hand, evaluated based on whether a warned property was or not subsequently flooded. They are explained as follows.

5.4.1 Flooding threshold-based criterion (FTC)

A flood in an at-risk community is defined as the overtopping of the flooding threshold y_T, which is the river level at which the flood impacts begin (Figure 5.2a). This value is set at T_{SoP}, i.e., the return period that defines SoP, the Standard of Protection, which is an input parameter of the MC framework. As was mentioned above, the FTC evaluates the performance of a FEWS based on whether the occurrence or non-occurrence of a flood event in the at-risk community was preceded by a warning. In other words, both the warning decision and the continuous value of y defined by $f(y|\hat{y})$ are viewed categorically (or binary for 'yes'/ 'no' decisions and events respectively) according to whether a value of y that exceeded or not the flooding threshold y_T was preceded by a warning. The two-by-two contingency table shown in Table 5.2 was used to describe the resulting four potential outcomes one can find under this criterion. They are defined as: a hit, if the flood event occurred and a warning was issued (h describes the total number of hits); a miss, if the flood event occurred and a warning was not issued (m describes the total number of misses); a false alarm, if the flood event did not occur and a warning was issued (f describes the total number of false alarms); and a correct negative, if the flood event did not occur and a warning was not issued (cn describes the total number of correct negatives). The sum of the total number of occurrences (e) and non-occurrences (e') of a flood event define the total number of events analysed (n). The total number of warnings is denoted as w, and the total number of no-warnings as w'.

Table 5.2 Two-by-two contingency table to analyse the performance of a FEWS based on the flooding threshold-based criterion FTC.

Observations	Warning	No Warning	Total
Flood event	h	m	e
Non-flood event	f	cn	e'
Total	w	w'	n

5.4.2 Floodplain property-based criterion (FPC)

As was explained in section 5.2, the FPC can be used in FEWSs that consider the forecast of the flood extent to warn vulnerable properties. In this case, the warning procedure can first include a 'yes'/ 'no' warning decision based on the river level forecasts and the flooding threshold y_T, and, then, a dissemination process can be

done based on the inundation forecasting. Thus, the performance of this type of FEWSs depends on the warning decision and the uncertainty of the inundation forecast, and its analysis can be based on whether the warned properties were or were not subsequently flooded. The target information for this analysis is, therefore, the number of houses warned n_{houses}^{warned} and the number of houses flooded $n_{houses}^{flooded}$ which, as was explained in section 5.3, in this work, is represented by an impact-curve-based method. A two-by-two contingency table (Table 5.3) can also be built based on this information. In this case, the four potential outcomes after the occurrence or non-occurrence of a flood event are defined as: a hit, if a house is warned and flooded (h describes the total number of hits); a miss, if a house is not warned and flooded (m describes the total number of misses), a false alarm, if a house was warned and not flooded (f describes the total number of false alarms); and correct negative, if a house is not warned and not flooded (cn describes the total number of correct negatives). The sum of the total number of houses flooded and not flooded defines the total number of houses analysed.

Table 5.3 Two-by-two contingency table to analyse the performance of a FEWS based on the floodplain property-based criterion FPC.

Observations	Warning	No Warning	Total
House flooded	h	m	Total no. houses flooded
House not flooded	f	cn	Total no. of no flooded houses
Total	Total no. houses warned	Total no. of no warned houses	Total no. houses analysed

Figure 5.4 illustrates how the warning decision and uncertainty of the inundation forecasting affect the performance of this type of FEWS. Figure 5.4a shows the case when a flood was preceded by a warning. Note that this would represent a hit in the FTC; in the FPC, however, depending on the magnitude of the forecast and observed flood event, i.e., y and $\hat{y}$, one can find misses or false alarms. Figure 5.4b and Figure 5.4c show the case when a wrong warning decision was made. In this case, misses are due to the fact that the warning dissemination was not conducted before a flood occurred (Figure 5.4b), and false alarms (Figure 5.4c) due to the fact that the dissemination was conducted when a flood did not occur.

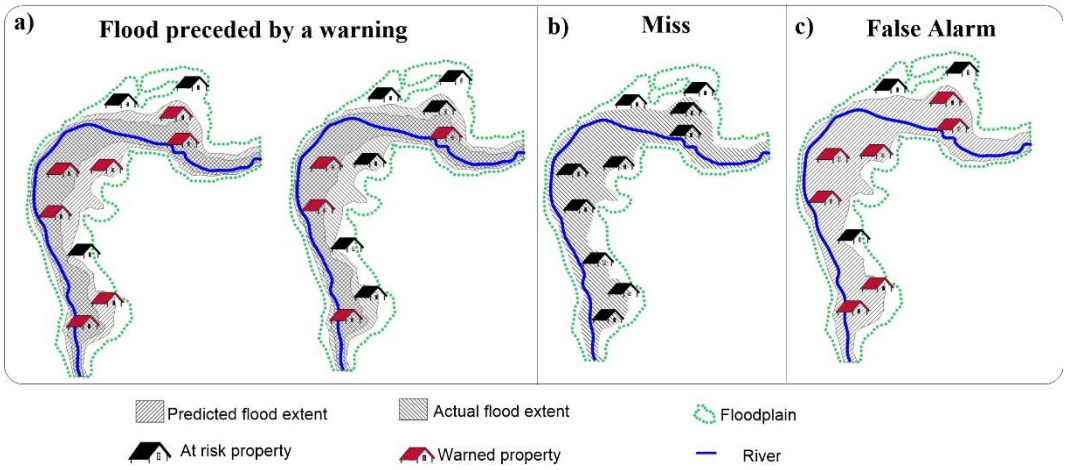


Figure 5.4: Illustration of the potential outcomes in the floodplain property-based criterion FPC.

5.5 Metrics for evaluating the flood warning performance

As was mentioned in sections 5.4 and 2.7.1, the probability of detection POD and the false alarm ratio FAR have been defined by Parker (2017) as two measures of the reliability of a FEWS. In this work, two of these measures have been chosen to describe this important characteristic of the flood warning performance. These metrics are computed based on the two-by-two contingency tables explained in the prior section. Table 2.2 gives a description of them, which are briefly described as follows.

POD is also known as hit rate and provides an estimate of the probability that an observed event being warned. It ranges from 0 to 1, and a perfect forecast is 1. It is computed as the ratio between the total number of hits h and the total number of observed events (h+m). As it is based only on observed events, it is only sensitive to miss events and not to false alarms. This skill score is, therefore, incomplete by itself and should be used in conjunction with, for example, FAR that provides an estimate of the probability that a forecast event being incorrectly warned. It ranges from 0 to 1, with 0 representing a perfect forecast. It is computed as the ratio between the total number of false alarms f and the total number of forecast events (h+f). This skill score is also incomplete, as it is not sensitive to miss events. POD can be artificially inflated by issuing more warnings, and FAR can be artificially reduced by issuing fewer warnings; hence these metrics are used together

In the FTC, the POD and FAR answer the following questions: What is the probability of an observed event being correctly warned? and What is the probability of a forecast event being incorrectly warned? respectively. In the FPC, the POD and FAR answer the following questions: What is the probability of a flooded house being correctly warned? What is the probability of a warned house being incorrectly warned? respectively.

5.6 Flood warning decision rule approaches

As was explained in section 5.4, the contingency table-based method used in the MC framework to define the reliability of a FEWS depends on the warning decisions. This section of the chapter describes how these decisions are represented in the MC framework in the FTC and FPC. Deterministic forecast-based and probabilistic forecast-based decisions are considered in the analysis. They are represented through a decision rule approach and are described as follows.

5.6.1 The deterministic forecast-based decision rule

The notation used for this rule is DFDR. This decision rule assumes that the warning decision is driven by the deterministic forecast of the flood magnitude, i.e., $\hat{y}$. For the FTC and FPC, the DFDR assumes that a warning is issued when $\hat{y}$ is greater than the flooding threshold y_T.

$$\begin{aligned} &if\ (\hat{y} > y_T)\ \ warn, \\ &else, \qquad not\ warn \end{aligned} \qquad \text{Eq. 5.1}$$

For the FPC, the number of warned properties n_{houses}^{warned} associated with a value of $\hat{y}$ is derived from the impact curve of the at-risk community (Figure 5.3b). Figure 5.5 gives an illustration of the DFDR for the FTC and FPC, whereas Table 5.4 describes them.

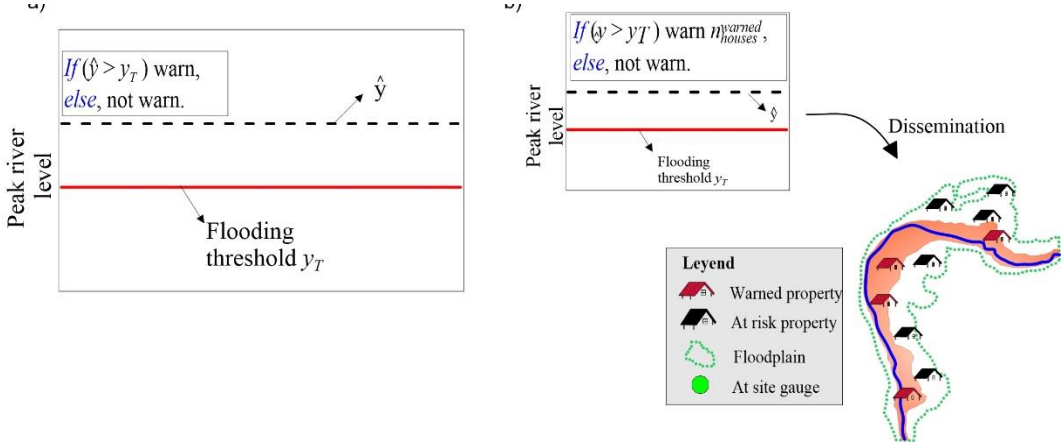


Figure 5.5: Illustration of the deterministic forecast-based decision rule (DFDR) for a pair $(y, \hat{y})$ for the flooding threshold-based criterion FTC (a) and floodplain property-based criterion FPC (b).

Table 5.4 Description of the deterministic forecast-based decision rule (DFDR) for the flooding threshold-based criterion FTC and floodplain property-based criterion FPC.

Decision rule	Abbreviation	Criterion	Decision rule
Deterministic forecast-based decision rule	DFDR	FTC	*If* ($\hat{y}$>y_T) warn, *else*, not warn
		FPC	*If* ($\hat{y}$>y_T) warn n_{houses}^{warned}, *else*, not warn

5.6.2 The probabilistic-based decision rules

Two types of probabilistic decision rules were considered. One was optimized based on the FTC, and the other based on the FPC. The notation used for the former is PTDR, which assumes that the Warner makes a decision based on a probabilistic threshold (PT) and the probability of exceedance (PE) of the flooding threshold y_T. i.e., the probability of having an observed river peak level y greater or equal than y_T. The probability of exceedance PE is a value delivered by the MCFG which is obtained through the conditional distribution of y given $\hat{y}$, i.e., $f(y|\hat{y})$, derived from the joint probability of these two variates $f(y, \hat{y})$ that is assumed here to be a bivariate Lognormal distribution. The decision rule, in this case, is given by:

$$\begin{aligned} &if\ (PE > PT)\ \ warn, \\ &else, \qquad not\ warn \end{aligned} \qquad \text{Eq. 5.2}$$

The notation of the other probabilistic rule is PDR. It also uses a probabilistic threshold PT. However, in this case, PT defines the warning level $\hat{y}_w$ from $f(y|\hat{y})$

at which the Warner should base his/her decisions. Then, the warning decision can be based on the direct comparison between $\hat{y}_w$ and the flooding threshold y_T as:

$$\begin{aligned} &if\ (\hat{y}_w > y_T)\ \ warn\ n_{houses}^{warned}, \\ &else, \qquad not\ warn \end{aligned} \qquad \text{Eq. 5.3}$$

The aim of the PTDR and PDR is, thus, to find an optimal probabilistic threshold PT^* that maximises the difference between POD and FAR based on the FTC and FPC, respectively. Figure 5.6 illustrates the application of PTDR and PDR for the FTC and FPC, respectively, whereas Table 5.5 gives a description of them.

Table 5.5 Description of the probabilistic threshold-based decision rule (PTDR) for the flooding threshold-based and floodplain property-based criterion, FTC and FPC, respectively.

Decision rule	Abbreviation	Criterion	Decision rule	Description
Probabilistic threshold-based decision rules	PTDR	FTC	*If* (PE>PT) warn, *else*, not warn	PT optimized based on FTC
	PDR	FPC	*If* ($\hat{y}_w$>y_T) warn n_{houses}^{warned}, *else*, not warn	PT optimized based on FPC

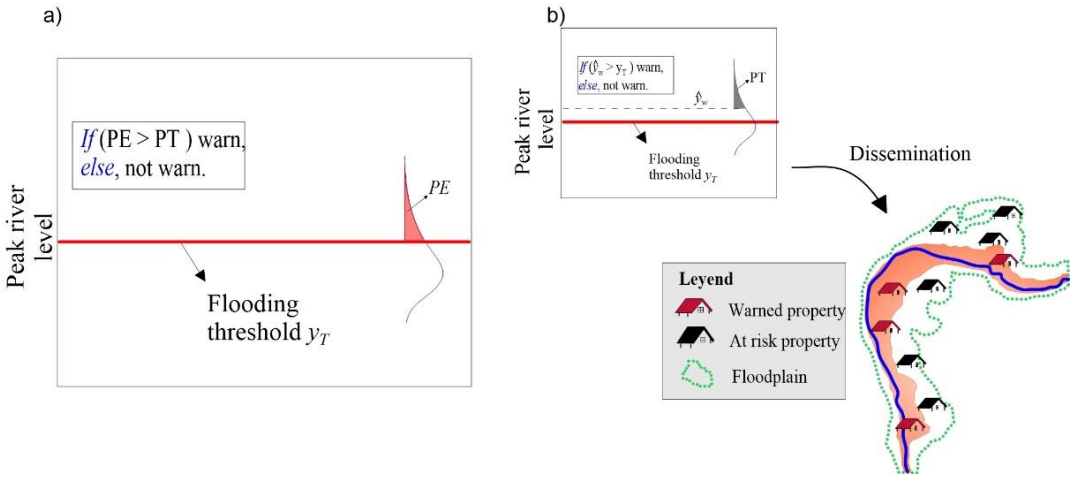


Figure 5.6: Illustration of the probabilistic rules PTDR and PDR for flooding threshold-based criterion FTC (a) and floodplain property-based criterion FPC (b), respectively.

5.7 Sensitivity experiments

In the prior sections, the criteria, metrics, and warning decision rules considered in the MC framework to explore the reliability of a FEWS were introduced. As was mentioned in the introductory part, the analysis of this characteristic of the FEWS

requires the addition of two parameters to the ones considered in the MCFG. In this last section of the chapter, the sensitivity of FEWS performance to several of these parameters that represent the main assumptions in the estimation of the reliability of a FEWS is explored. This is done through several sensitivity experiments based on the one-at-a-time (OAT) method (Pianosi *et al.*, 2016), which, in essence, varies/perturbs the input parameters of the MC framework from its reference parameter values (baseline scenario) one at a time and assesses the impacts on the metrics used to define the reliability of a FEWS. Thus, in this section, the parameters and reliability that define the baseline scenario are first described, and then, the results of the sensitivity analysis are shown.

5.7.1 Baseline Case

The key parameter controlling the performance of a flood forecasting model is the correlation between the forecast and observed values for a specified lead time τ. There is a large and varied literature on the performance of rainfall-runoff models when used in simulation mode; this is typically measured by the Nash-Sutcliffe Efficiency (NSE), which, for an unbiased model, equals ρ^2. Values of NSE are computed from continuous simulations and are dependent on the time unit. The most widely quoted values in the literature are for daily streamflow simulations; for example, a study where TOPMODEL was calibrated for a set of 1013 UK catchments reported a range of NSE values ranging from -2.47 to 0.94 with a median of 0.77(Lane *et al.*, 2019). Similar studies for hourly streamflow are not reported due to the lack of long records of hourly rainfall data. Views of what is an acceptable value of NSE vary widely and depend on the intended application, but typically, a value greater than 0.80 might be considered acceptable. For hourly streamflow, there are no definitive findings on whether the NSE values might be larger or smaller than for daily values. Consequently, if just the peaks are sampled from continuous simulations, the NSE can be expected to be lower than for the continuous simulation case where the smaller errors in recession periods will increase the correlation.

In the case of a real-time forecasting model, forecast updating will lead to much higher NSE values for short lead times; as the lead-time τ increases up to the catchment lag L, the effect of updating will die out, and the NSE values will decline towards the simulation mode value. The selected time unit will be a function of the

catchment lag L. Values of NSE for real-time flood forecasts are rarely quoted by agencies responsible for operational flood forecasting. A research study for the Eden at Sheepmount (Leedal *et al.*, 2013) quoted a value of NSE = 0.98 for six hour ahead forecasts of water levels (using hourly data, with updating) for levels less than 2m, and 0.45 for values greater than 2m; the latter figure represents only a small percentage of the data but highlights the problem of declining performance in forecasting large floods. It should be noted that the bivariate sampling of forecast and observed flood levels from the BLND represents this very well, as the forecasting errors increase with level.

Taking the above points into account, it was decided to adopt a correlation of 0.85 for the peak levels in the baseline case, which, for an unbiased model, corresponds to an NSE of 0.72; the corresponding NSE for continuous simulation would be higher. This correlation was taken to correspond to a catchment lag L and lead time τ of 6 hours, i.e., the IP_ρ value in the lead time-correlation function (Figure 4.14). The aim then is first to find where in the POD/FAR performance space this model is located and then to explore how perturbations to the set of chosen baseline parameters, and the correlation, in particular, can enhance or degrade POD/FAR performance. In particular, the aim is to identify the levels of correlation and NSE needed to achieve different levels of reliability for flood warnings, as measured by POD and FAR.

The baseline case assumes that there are 1000 houses at-risk in the floodplain, and it has 1 in 5-year SoP (T_{SOP}=5years). When the water level overtops the SoP, houses are assumed flooded from the bankfull height upwards (Figure 5.7). MC simulations were performed using 30 replications, where each replication consisted of 10,000 events. Finally, the MC estimates were obtained from the average values of 30 replications, where each replication consisted of 10,000 simulated events.

Table 5.6 depicts the input parameters of the MC framework and the values which define the baseline scenario. They are split according to the component they represent. The baseline model assumes an ideal case of the forecasting system where the marginal distribution of the observations is equal to that of their forecasts; therefore, the moments of $\hat{y}$ are equal to those of y. The moments of y have been adopted by taking as reference the parameters obtained for the station River Wansbeck at Mitford (Chapter 4).

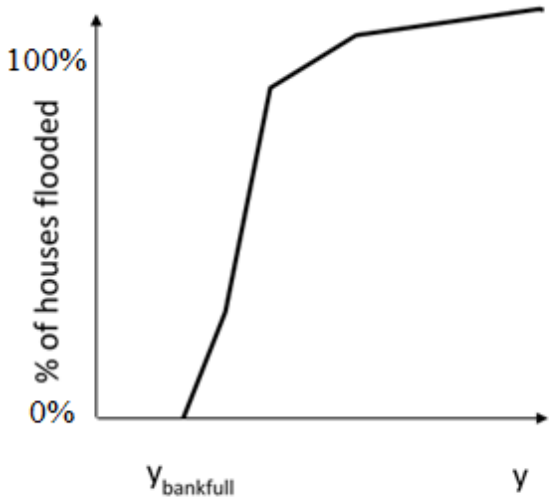


Figure 5.7: The impact curve used to quantify the number of flooded houses.

Table 5.6: Assumed input parameter values for the baseline case

Abbreviation	Description	Value adopted	Parameter associated with:
μ_y	Mean of y.	2.51	The river basin
σ_y^2	The variance of y.	0.20	
y_o	Location parameter of y	2.03	
γ_y	The average number of peaks per year	1.60	
L	Basin lagtime L	6 hrs	
$\mu_{\hat{y}}$	Mean of the forecasts of y	μ_y	Flood forecasting system
$\sigma_{\hat{y}}^2$	Variance of the forecasts of y	σ_y^2	
$\hat{y}_o$	Location parameter of the forecasts of y	y_o	
$\gamma_{\hat{y}}$	The average number of peaks per year of the forecasts of y	γ_y	
IP_ρ	The inflection point of the lead time-performance function	0.85	
τ	Lead time	6 hr	
T_{ToS}	Return period associated with the flooding threshold y_T.	5 years	The at-risk community
$n_{houses}^{at-risk}$	Total number of houses at risk in the benefit area	1000	
n	Sample size or number of simulated events (30 replications)	10000	Monte Carlo simulation

Figure 5.8a shows the joint density of 10,000 pairs $(y, \hat{y})$ generated from the MCFG which, as was explained above, uses a bivariate Lognormal distribution, BLND, to represent the bivariate values. Since the moments of $\hat{y}$ are equal to those of y, the generated bivariate values represent an ideal forecast scenario where the statistics of the observations are equal to their forecasts. This bivariate simulation was done

based on the algorithm shown in section 4.4.4. Figure 5.8b shows an example of the PU expressed as $f(y|\hat{y})$ for a forecast $\hat{y}$, and the probability of the flooding threshold being exceeded.

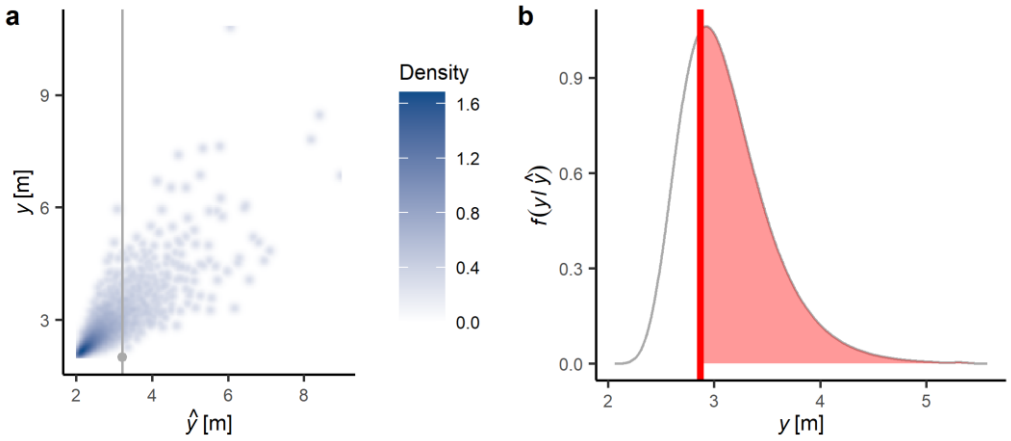


Figure 5.8: Joint density of pairs (y,$\hat{y}$) generated from the Monte Carlo flood and forecast generator (MCFG) assuming a bivariate lognormal distribution (BLND).
Figure a shows the joint density of 10,000 bivariate pairs ($y, \hat{y}$) generated from the MCFG assuming a BLND defined by the values of moments of y and $\hat{y}$ assumed in the baseline scenario. Figure b shows the conditional distribution $f(y|\hat{y})$ associated with the forecast value of $\hat{y}$ represented by a grey dot in Figure a. $f(y|\hat{y})$ is obtained by slicing the joint distribution through a grey line that crosses the forecast value (grey dot). Knowing this predictive density, one can estimate the probability of exceedance (PE) of the flooding level (red area), which is the probability value considered in the PTDR. The flooding level is defined by the assumed SoP in the baseline scenario (5 years).

The analysis of the performance of the baseline FEWS under the above-mentioned criteria is summarised in Figure 5.9. This figure shows the optimization procedure to find the optimal probabilistic threshold PT^* for the PTDR and PDR. In Figure 5.9a and Figure 5.9b, the blue and grey lines indicate how the values of POD and FAR change when considering several values of the probabilistic threshold PT for the PTDR and PDR, respectively. These values are summarised in the FAR-POD curve shown in c. A point on these lines is associated with a value of PT. The values associated with the optimal probabilistic threshold PT^* are represented with a square and circle for the PTDR and PDR, respectively. The figure also shows the results of the DFDR for the FTC and FPC. They represent a point on the FAR-POD curve and are represented with a triangle and diamond, respectively. The main conclusions of this figure are explained as follows.

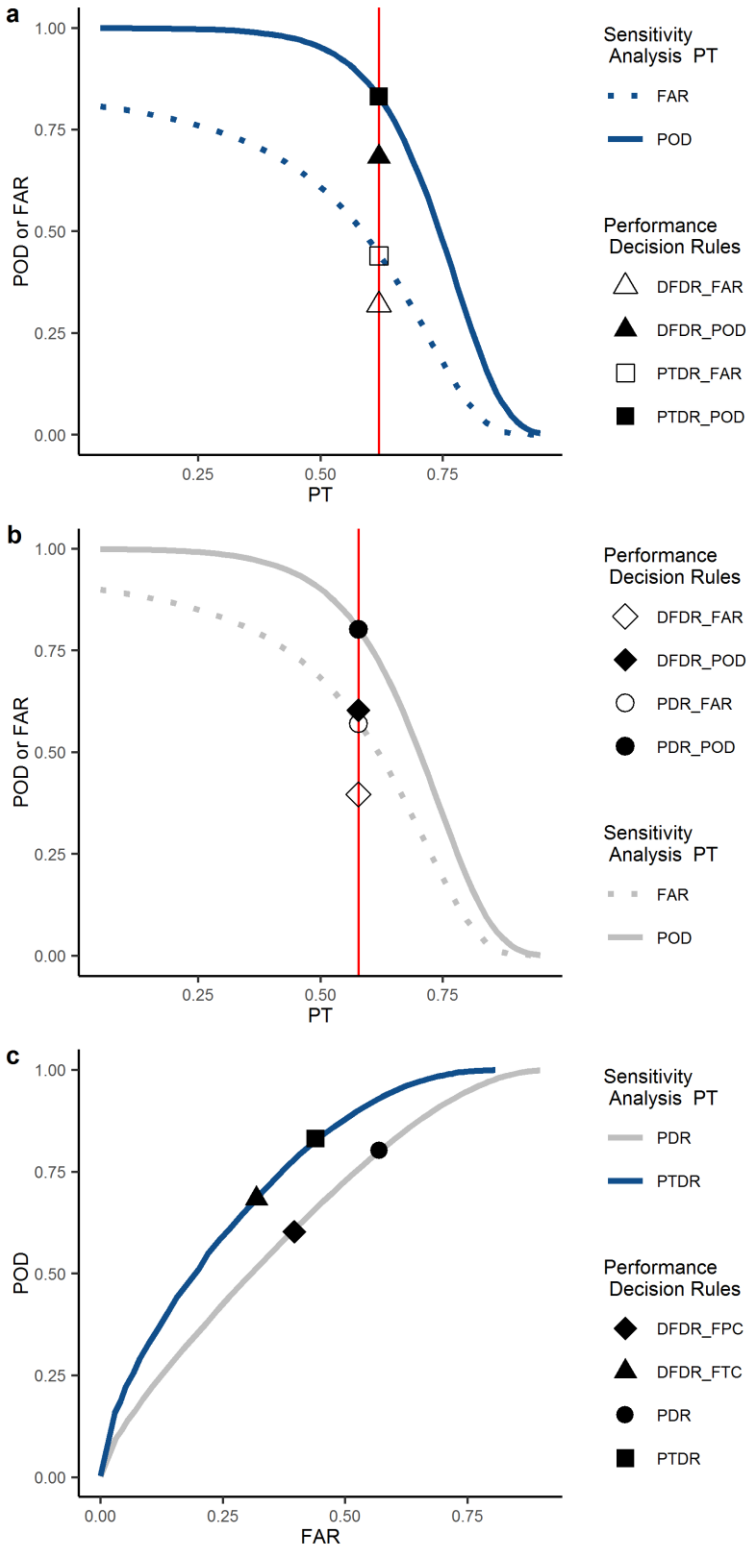


Figure 5.9: Analysis of the performance of the baseline FEWS through the POD-FAR curves for the flooding threshold-based and floodplain-property-based criterion, FTC and FPC, respectively.

This figure shows the FAR and POD results for different values of PT for the PTDR (a) and PDR (b). These values are summarised in the FAR-POD curve shown in c. These decision rules' optimal values are represented with a square (PTDR) and circle (PDR). A triangle and diamond are used to show the results of the DFDR for the FTC and FPC, respectively.

POD: Probability of detection; FAR: false alarm ratio; PT: probabilistic threshold, DFDR: deterministic forecast-based decision rule (DFDR), PTDR: probabilistic rule based on a PT and the probability of exceedance (PE) of y_T, PDR: probabilistic rule based on a PT and the PE of y_w.

The performance of the baseline FEWS based on the probabilistic decision rules PTDR and PDR can be regarded as better than that obtained with the deterministic decision rule DFDR for the FTC and FPC, respectively, because the associated difference between POD and FAR is higher (Table 5.7, column 8). The results also show that the POD and FAR values associated with the probabilistic rules are higher than those obtained with DFDR(Table 5.7, columns 6 and 7). That means that the former decision rules tend to increase the value of POD, which necessarily increases FAR. Thus, even though the flood warning performance associated with PTDR and PDR is better than its deterministic version, the probabilistic-decision-rules- based results tend to deliver relatively high values of FAR, 0.43 and 0.55 for PTDR and PDR, respectively (Table 5.7). These values are higher than those obtained with DFDR, and it is due to the fact that the optimization criterion for PTDR and PDR does not place any restriction on the value of FAR, and it looks for the biggest difference between POD and FAR, regardless of the value of the latter. It can be understood as the price being paid for obtaining a high value of POD, which is considerably higher than for the deterministic case (Table 5.7). Thus, the optimization criterion adopted in this research (maximize POD minus FAR) gives higher weight to the reduction of missed events whose economic consequences, in terms of floods, far outweigh those associated with false alarms. The effect of having high values of FAR in a warning system is known as 'cry wolf', which has to do with the disregarding of flood warnings due to their loss of credibility as a result of the high percentage of false alarms. Target values of FAR are usually between 0.2 and 0.5 (Jolliffe and Stephenson, 2012). However, Barnes *et al.* (2007) advocate that there is little evidence that a high value of FAR causes users to ignore warnings of severe events.

Table 5.7: Contingency table based on the flooding threshold-based and floodplain property-based criterion, FTC and FPC, respectively, for the baseline FEWS for deterministic and probabilistic decision rules and optimal probabilistic thresholds.

Lead time (τ) [h]	Prob. Thres.	h	m	f	POD	FAR	POD-FAR	Decision rule	Criterion
6	-	858	395	389	0.68	0.31	0.36	DFDR	FTC
	-	170315	110467	110668	0.61	0.39	0.20		FPC
	0.62	1041	212	812	0.83	0.43	0.39	PTDR	FTC
	0.59	220819	59962	283208	0.79	0.55	0.23	PDR	FPC

An interesting point in the analysis is to see how the performance of a FEWS obtained based on FPC is related to the performance one would obtain if it is quantified in terms of FTC. These results can be analysed directly from the values obtained based on FTC and FPC for the DFDR (first and second row in Table 5.7). The results show that the performance based on the FPC is lower than that obtained based on the FTC. That difference is directly related to the uncertainty of the flood magnitude, i.e., the difference between y and $\hat{y}$ which define the difference between the warned and flooded properties. Note that a hit in the FTC does not mean a hit in the FPC; due to the difference between these two variables, a hit in the former might include misses or false alarms in the latter (Figure 5.4a). This implies that if inundation level forecasting is undertaken (FPC), then the performance in terms of POD and FAR will drop relative to the FTC, which implies improved forecasts would be needed to achieve the same performance level as for FTC.

5.7.2 Sensitivity experiments

In this last sub-section of the chapter, the sensitivity of several parameters with respect to the results shown in Table 5.7 is explored. Three sensitivity experiments. have been carried out.

5.7.2.1 Sensitivity experiment 1

It is well known that the forecast performance decreases as the lead time increases (Blöschl, 2008; Parker, 2017). The forecast performance in the joint distribution is controlled by the moments of $\hat{y}$ and the correlation coefficient $\rho_{y\hat{y}}$. The latter is assigned to a lead time τ through the lead time-correlation function shown in Figure 4.14. As has been explained, τ is an input parameter of the framework. One can assume that $\hat{y}$ are outputs of different flood forecasting systems associated with different lead times τ and that forecasts are represented by the same moments. Under this assumption, it is hypothesized that the forecast performance is mainly controlled by $\rho_{y\hat{y}}$ (a bias in the mean or variance of the forecasts can also affect performance) which, in turns, depends on the input value of τ. This first experiment is based on this assumption and aims to analyse how the outcomes of the baseline scenario change when considering different forecast performances represented by different values of τ and assuming the same moments of $\hat{y}$ for each of them. Table 5.8 gives a description of this experiment and the range of values to be explored.

Table 5.8: Description of sensitivity experiment 1

Parameter to be modified	Lower bound	Baseline	Upper bound	Aim
τ	3 hr (-50%)	6 hr	24 hr (+200%)	Analyse how the flood warning performances obtained in the baseline scenario changes as the lead time changes.

The results of the sensitivity experiment are described in Figure 5.10, which shows two diagram types that show, for each value of τ, its associated pair of values of POD and FAR obtained from the FTC and FPC for DFDR, PTDR, and PDR. Points on the blue line are skill scores computed based on FTC, and points on the grey line are those computed by considering FPC. One of the diagram types (a and c) shows POD and FAR on the left axis and lead-time on the abscissa, whereas the other, a quadrant plot (b and d), shows these values on the POD/FAR performance space. In the quadrant plots, the value of POD and FAR equal to 0.5 have also been added as a reference. These values define four quadrants: the left-upper quadrant with values of POD > 0.5 and FAR < 0.5; right-upper quadrant with values of POD > 0.5 and FAR > 0.5; left-bottom quadrant with values of POD < 0.5 and FAR < 0.5; and right-bottom quadrant with values of POD < 0.5 and FAR > 0.5. The 1:1 line (grey dashed line) also defines points with values of POD greater than FAR (points above this line) or the opposite (points below this line). Three target levels of reliability are defined here based on POD and FAR:

Low: POD > 0.5, FAR < 0.5

Medium: POD > 0.7, FAR < 0.3

High: POD > 0.8, FAR < 0.2

These levels represent a simplified form of that used in England for the national assessment of flood forecasting systems(Robson *et al.*, 2017). The results of the sensitivity experiments can be judged against these target levels. They are explained as follows.

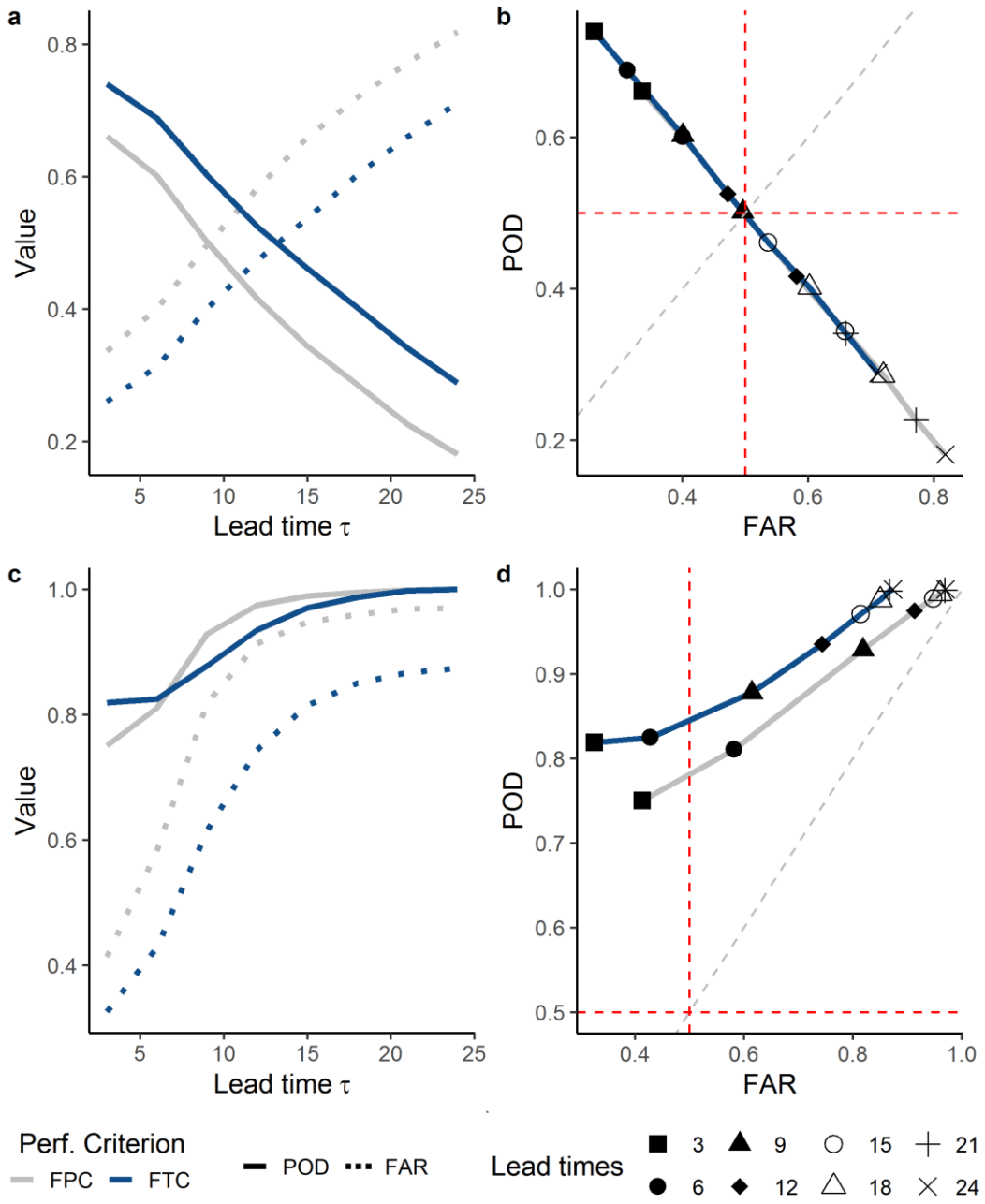


Figure 5.10: Results of the sensitivity experiment 1 for the flooding threshold-based and floodplain property-based criterion, FTC and FPC, respectively, based on the deterministic rule DFDR and probabilistic rules PTDR and PDR.

Figure a shows FAR and POD results for different lead times τ based on the DFDR, whereas Figure b shows these values for the PTDR and PDR. FTC-based results (blue line) are shown for the DFDR and PTDR, and FPC-based results (grey line) are shown for the DFDR and PDR. The values of these figures are summarised in the FAR-POD curves shown in b and d, respectively. *POD: Probability of detection; FAR: false alarm ratio; PT: probabilistic threshold, DFDR: deterministic forecast-based decision rule, PTDR: probabilistic rule based on a PT and the probability of exceedance (PE) of y_T, PDR: probabilistic rule based on a PT and PE of y_w, FTC: flooding threshold-based criterion, FPC: floodplain property-based criterion.*

Figure 5.10a and Figure 5.10b shows how POD and FAR change with lead time τ based on the deterministic decision rule DFDR for the FTC and FPC cases. These figures confirm, what was shown in the baseline case, that the POD and FAR values are lower for the FPC than the FTC. As was explained, it is due to the results based

on the FPC depend on the warning decision and the uncertainty of the inundation forecasting (Figure 5.4). The latter is represented in this framework as the difference between the warned and flooded properties which are obtained as a function of $\hat{y}$ and y respectively (Figure 5.3). In addition, as one can see in the quadrant plot, as the value of τ increases, the skill scores tend to move from the left upper quadrant to the right bottom quadrant, which is one of the worse scenarios of the system as the value of POD is lower than 0.5 and lower than FAR, which, in turn, is higher than 0.5. The lead time at which this occurs is approximately 13 hr ($\rho_{y\hat{y}}$=0.64) for the performance based on the FTC, and 10 hr ($\rho_{y\hat{y}}$=0.73) for the skills scores based on the FPC. FEWS with POD and FAR values in this quadrant can be considered as not operationally useful. In this sense, the results show that, for the FTC, deterministic FEWS can have medium (τ=3, $\rho_{y\hat{y}}$=0.89 , or τ=6, $\rho_{y\hat{y}}$=0.85) or low reliability (τ=9, $\rho_{y\hat{y}}$=0.76, or τ=12, $\rho_{y\hat{y}}$=0.67). However, if one considers the FPC, deterministic FEWS can have low reliability (τ=3, $\rho_{y\hat{y}}$=0.89, or τ=6, $\rho_{y\hat{y}}$=0.85).

Figure 5.10c and Figure 5.10d show how POD and FAR change with the lead time τ based on the probabilistic decision rules PTDR (based on the FTC) and PDR (based on the FPC). These figures again confirm that the flood warning reliability based on the FPC is worse than that based on the FTC. In addition, the quadrant plot shows that as the value of τ increases, the skill scores tend to move from the left upper quadrant to the right upper quadrant. This occurs because, as was mentioned above, these probabilistic-based decision rules tend to deliver high values of POD and FAR due to the fact their optimization criterion does not place any restriction on the value of FAR, and it looks for the biggest difference between POD and FAR, regardless of the value of the latter. The results show that this drawback of the optimization criterion makes the FAR value increases as the lead time τ increases. The right upper quadrant is the other worse scenario of the system as the value of FAR is higher than 0.5. The lead time at which this occurs is approximately 8 hr ($\rho_{y\hat{y}}$=0.79) for the performance based on the PTDR, and approximatly 5 hr ($\rho_{y\hat{y}}$=0.87) for the skills scores based on the PDR. FEWS with POD and FAR values in this quadrant can also be considered as not operationally useful. In this sense, the results show that probabilistic FEWSs based on the PTDR can have medium (τ=3 hr, $\rho_{y\hat{y}}$=0.89) or low reliability (τ=6 hr, $\rho_{y\hat{y}}$=0.85). However, if one considers the probabilistic decision rule PDR, a probabilistic

FEWS can have low reliability (τ=3 hr, $\rho_{y\hat{y}}$=0.89). These results suggest that the deterministic decision rule allows the FEWS to use longer lead times (lower $\rho_{y\hat{y}}$ values) with POD and FAR values operationally useful. The probabilistic decision rules tend to deliver high FAR values, as, as was mentioned above, they give higher weight to the reduction of missed events whose economic consequences, in terms of floods, far outweigh those associated with false alarms.

An interesting point in the DFDR-based results is to have an idea about the scatter of the pairs (y,$\hat{y}$) associated with the lead time at which the skill scores start to be in the right bottom quadrant in Figure 5.10b (13 hrs and 10 hrs approximately for the FTC and FPC, respectively). This allows us to have a good insight into how the bivariate values look for these poor performance cases. Figure 5.11 shows a sample of the bivariate values of the critical lead time for the FPC (10 hrs). These pairs (y,$\hat{y}$) are characterised by the moments of y and $\hat{y}$ used in the baseline scenario and the value of $\rho_{y\hat{y}}$=0.73 assigned to this critical lead time based on the lead time-correlation function shown in Figure 4.14

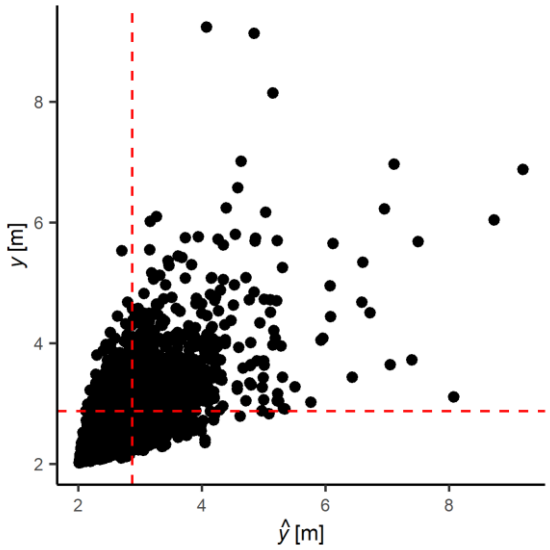


Figure 5.11: Scatter of the pairs (y,$\hat{y}$) associated with the critical lead time in the floodplain property-based criterion FPC based on the deterministic decision rule DFDR.
This figure illustrates a bivariate sample of the joint distribution that defines the critical lead time in the FPC based on the DFDR (10 hr) in the sensitivity experiment 1 (Figure 5.10b). The red dashed lines define the value of the flooding threshold y_T that define the hits, misses, false alarms, correct negative in DFDR in FTC. The value of $\rho_{y\hat{y}}$ for this joint distribution is 0.73.

5.7.2.2 Sensitivity experiment 2

This experiment explores the sensitivity to a change in the inflexion point IP_ρ of the lead time-correlation function shown in Figure 4.14 in conjunction with a change in lead time. Thus, the sensitivity of IP_ρ with respect to the results of the baseline scenario (Table 5.7) is analysed by modifying this parameter by a small percentage (± 10%). For a better interpretation of the results, this experiment also includes other values of τ, perturbed from the baseline case of 6 hours. Table 5.9 gives a description of this experiment and the range of values to be explored. In this case, several values within the range for τ were considered, whereas, for the IP_ρ, only the values of the lower and upper bound were included in the analysis.

Table 5.9: Description of sensitivity experiment 2

Parameter to be modified	Lower bound	Baseline	Upper bound	Aim
τ	3 hr (-50%)	6 hr	24 hr (+200%)	Analyse how the results of the baseline scenario change as the assumed forecast quality assigned to the lead time changes.
IP_ρ	0.77 (-10%)	0.85	.94 (+10%)	

The results of this sensitivity experiment are described in Figure 5.12 for the FTC and in Figure 5.13 for the FPC. These figures show how the values of POD and FAR associated with the baseline lead time τ (6hr) change when modifying IP_ρ and the lead time τ. The same diagram types used in the prior experiment are also used in this analysis. In the quadrant plots, points on the solid blue line are skill scores computed by considering the baseline value of IP_ρ (.85), points on the solid grey line are those computed by considering the lower bound of IP_ρ (.77), and points on the solid green line are those computed based on the upper bound of IP_ρ (.94), all for different values of τ. The black circle in Figure 5.12 and Figure 5.13 (6hr lead times) provides a reference point for showing how the results of the baseline scenario for the FTC and FPC (Table 5.7), respectively, change when modifying the value IP_ρ.

The upper plots in Figure 5.12 and Figure 5.13 show the results based on the deterministic decision rule DFDR for the FTC and FPC, respectively. Only the baseline lead time is considered here, and, as one can see, the results are very sensitive to changes in IP_ρ. For the upper bound value of IP_ρ, the values of POD

and FAR are improved considerably (higher values and lower values, respectively). For the lower bound value, they worsen to a lesser degree, but the change is still considerable. The same behaviour can be seen in the probabilistic forecast-based decisions, PTDR (bottom plots in Figure 5.12) and PDR (bottom plots Figure 5.13). To have a broader view of this behaviour, the analysis included other lead times. Note that this behaviour is the same if the baseline lead time would be 3hr (black square); however, if it would be 24h (black cross), the sensitivity of IP_ρ is reduced considerably. This occurs in larger degree in the results based on probabilistic forecast-based decisions, PTDR (bottom plots in Figure 5.12) and PDR (bottom plots Figure 5.13). These results bring an interesting conclusion; small improvements in the forecast quality (represented in this work as an increase of 10% of the value of IP_ρ (Table 5.9)) improve considerably the reliability of FEWS whose forecast quality is relatively good, the same benefits cannot be seen in FEWS whose forecast quality is relatively bad. Systems with good or bad forecast quality are presented in Figure 5.12 and Figure 5.13 as the points with small or long lead times whose values of $\rho_{y\hat{y}}$ is relatively high and low according to the lead time-correlation function shown in Figure 4.14, respectively.

The quadrant plots depicted in Figure 5.12 and Figure 5.13 also show further interesting points. Note that for the upper bound value of IP_ρ, more points are moved to the upper left quadrant which should be the target quadrant for a FEWS; this occurs for the deterministic and probabilistic scenarios for lead time beyond the assumed catchment lag L (6 hr). This shows how FEWSs improving the forecast quality can extend the lead time and, at the same time, have levels of POD and FAR that are operational useful. Furthermore, the quadrant plots for the probabilistic results (Figure 5.12d and Figure 5.13d) show different behaviour of the POD and FAR values for the FEWS with high correlation (upper bound value of IP_ρ). The results show that the probabilistic decision rules tends to decrease the FAR value until a given level of correlation $\rho_{y\hat{y}}$. This correlation value can be obtained by analysing the correlation associated with the lead time at which the FAR value start to increase. For the PTDR (Figure 5.12d) and PDR (Figure 5.13d), it occurs between a lead time of 6 and 9 hrs. According to the correlation-lead time function and the IP_ρ of this case (IP_ρ=0.94), the $\rho_{y\hat{y}}$ values associated with these lead times

are 0.94 and 0.85, respectively. The level of correlation $\rho_{y\hat{y}}$ until which the probabilistic results decrease the FAR value fall in this range.

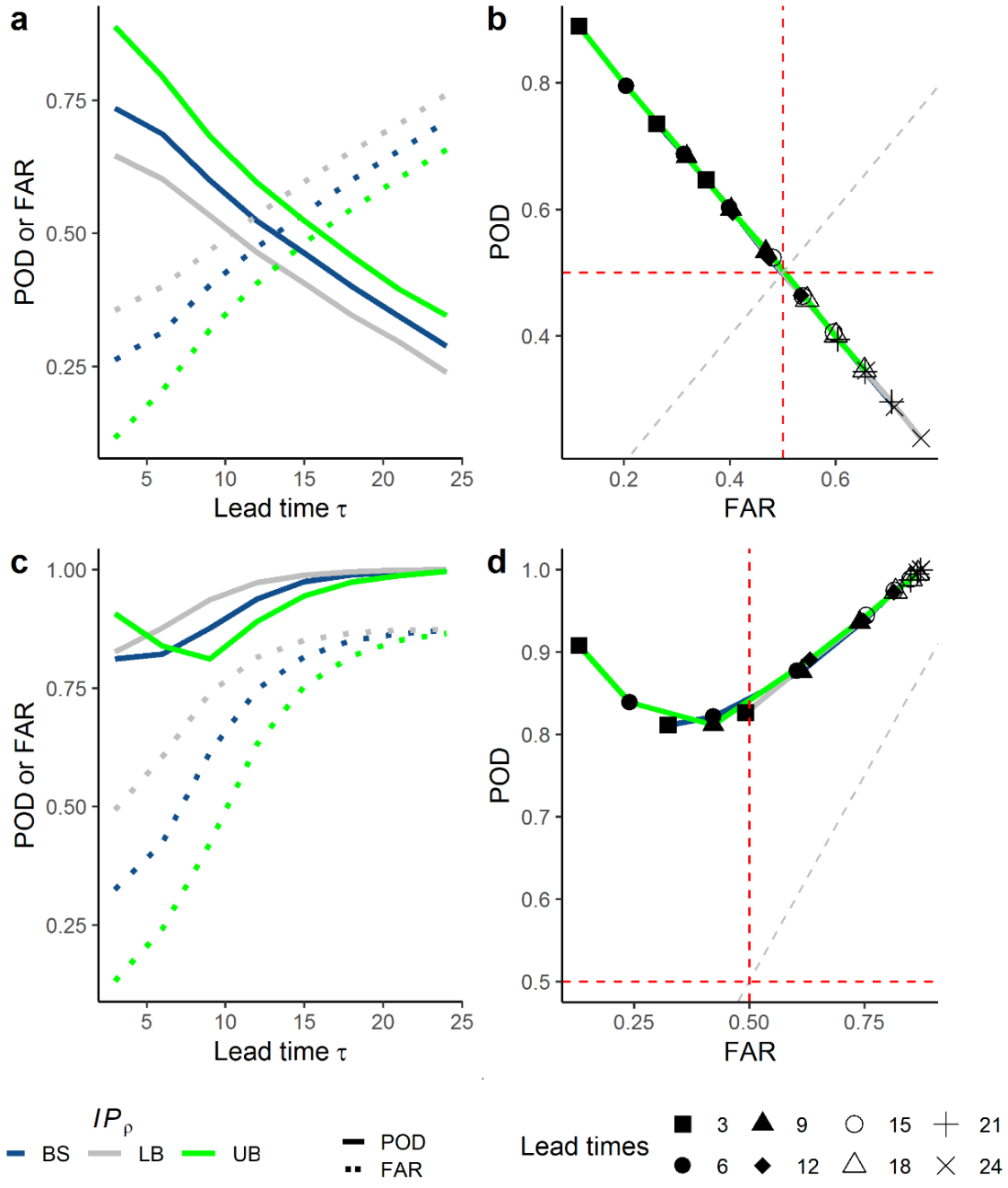


Figure 5.12: Results of the sensitivity experiment 2 based on the flooding threshold-based criterion FTC

Figures a and c shows FAR and POD results for different lead times τ based on the DFDR and PTDR and different values of IP_ρ, respectively. The values of these figures are summarised in the FAR-POD curves shown in b and d, respectively.

POD: Probability of detection; FAR: false alarm ratio; PT: probabilistic threshold, DFDR: deterministic forecast-based decision rule, PTDR: probabilistic rule based on a PT and the probability of exceedance (FTC) of y_T, BS: baseline value of IP_ρ, LB: the lower bound value of IP_ρ considered in the sensitivity analysis, UB: the upper bound value of IP_ρ considered in the sensitivity analysis. FTC: flooding threshold-based criterion

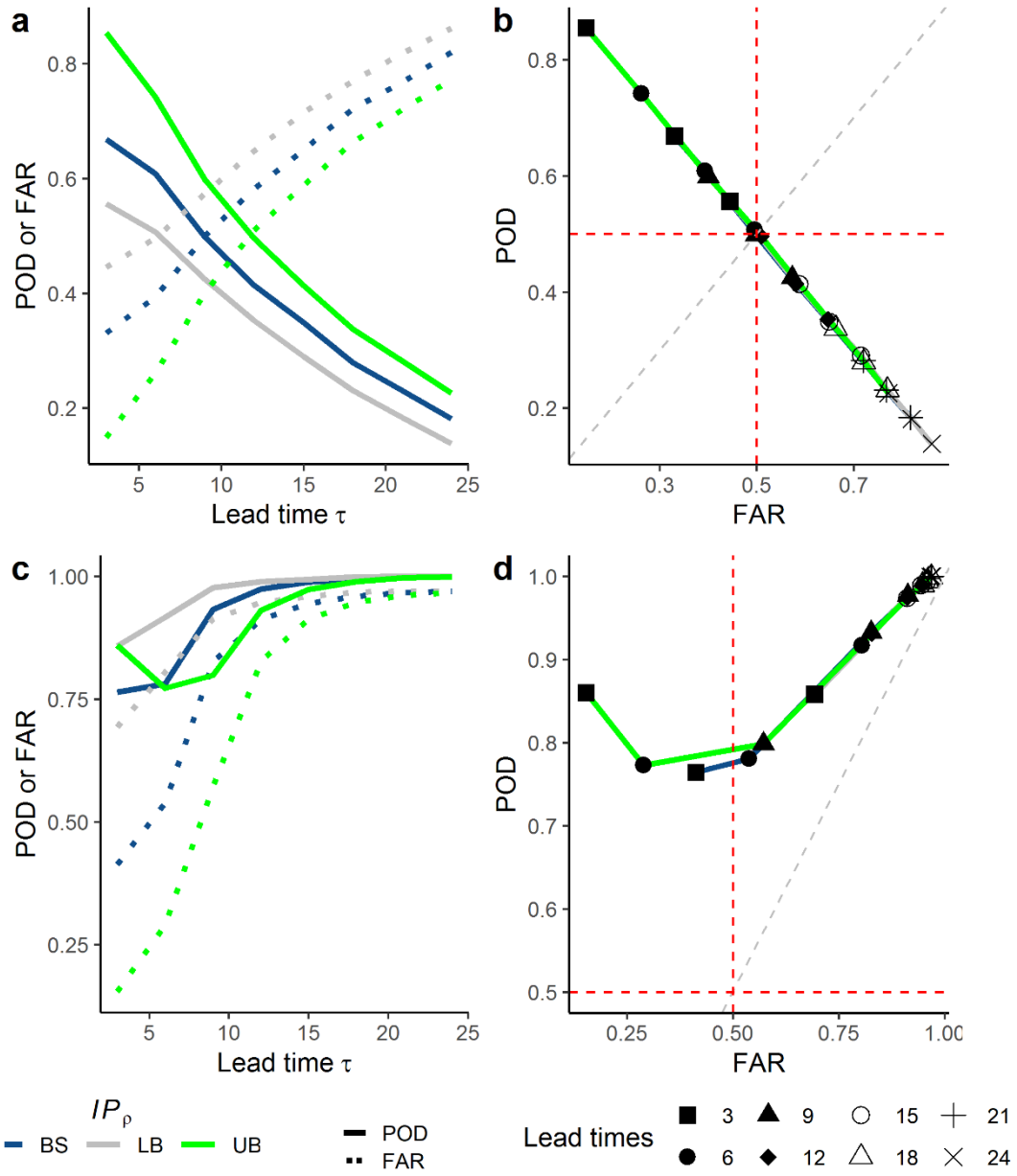


Figure 5.13: Results of the sensitivity experiment 2 based on the floodplain property-based criterion FPC

Figures a and c shows FAR and POD results for different lead times τ based on the DFDR and PDR and different values of IP_ρ, respectively. The values of these figures are summarised in the FAR-POD curves shown in b and d, respectively.

POD: Probability of detection; FAR: false alarm ratio; PT: probabilistic threshold, DFDR: deterministic forecast-based decision rule , PDR: probabilistic rule based on a PT and the probability of exceedance (PE) of y_w, BS: baseline value of IP_ρ, LB: the lower bound value of IP_ρ considered in the sensitivity analysis, UB: the upper bound value of IP_ρ considered in the sensitivity analysis. FPC: floodplain property-based criterion.

5.7.2.3 Sensitivity experiment 3

The sensitivity experiments 1 and 2 analysed the sensitivity of IP_ρ and $\rho_{y\hat{y}}$ with respect to the baseline results. These experiments, in essence, explored how the forecast quality can affect flood warning performance and assumed that the

moments of $\hat{y}$ are the same for each condition analysed. This last experiment assumes the opposite, i.e., IP_ρ and $\rho_{y\hat{y}}$ are assumed to be fixed values, and two moments of $\hat{y}$ ($\mu_{\hat{y}}$ and $\sigma_{\hat{y}}^2$) that also control the forecast quality are perturbed one at a time. The upper and lower bounds for the perturbations are shown in Table 5.10. The impacts of the perturbations on the marginal distribution of the forecasts relative to that of the observed values are shown schematically in Figure 5.14.

Table 5.10: Description of sensitivity experiment 3

Parameter to be modified	Lower bound	Baseline	Upper bound	Aim
$\mu_{\hat{y}}$	0.36 (-25%)	0.48	0.6 (+25%)	Analyse the bias in the mean and variance for the baseline case
$\sigma_{\hat{y}}^2$	0.10 (-50%)	0.20	0.30 (+50%)	

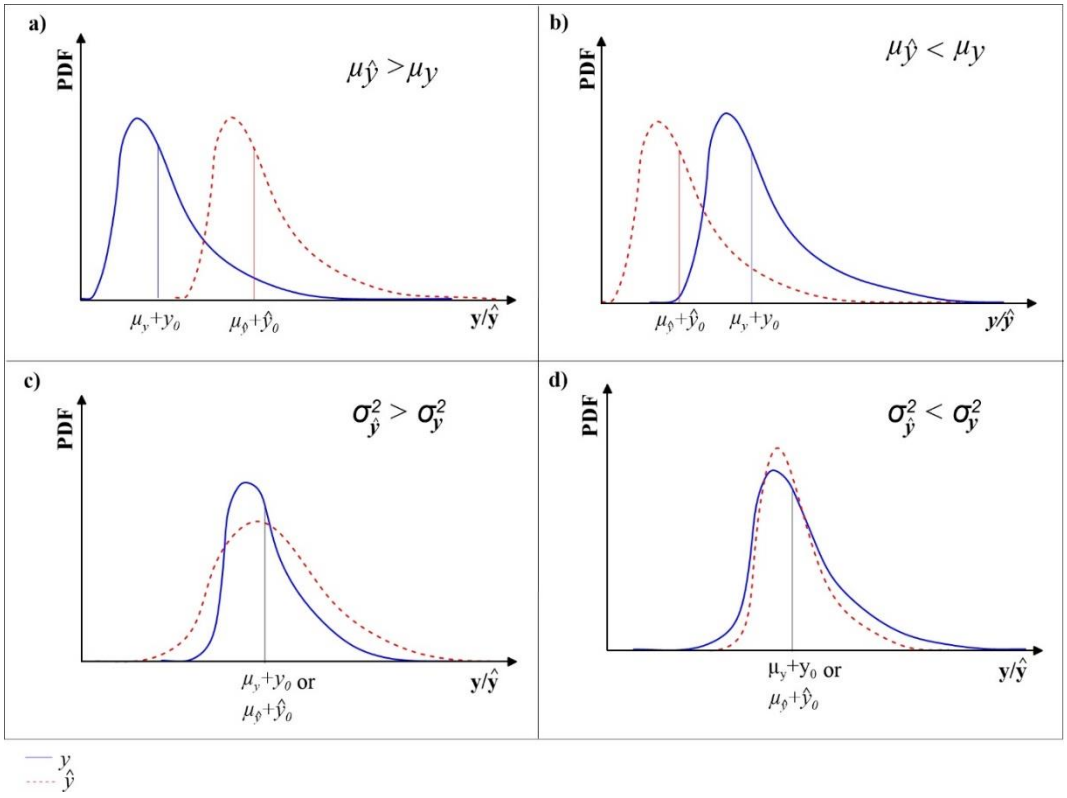


Figure 5.14: Illustration of how the marginal distribution of the forecast is perturbed in Experiment 3

The left and right upper and diagrams illustrate how the variations to $\mu_{\hat{y}}$ and $\sigma_{\hat{y}}^2$ perturbs the marginal distribution of $\hat{y}$ with respect to the marginal distribution of y, respectively.

It is well known that most forecasting models suffer from biases in the mean and variance, so the aim here is to provide insight into the impacts of these biases on POD and FAR for the deterministic and probabilistic forecasting cases. While a bias in the mean can be corrected, it is much less obvious what to do about a bias

in the variance. For all models which are fitted using least squares, there will be a loss of variance, expressed by the Nash-Sutcliffe Efficiency (NSE). In the case of forecasts of peak levels, a value of NSE = 0.80 implies that there is a 20% loss of variance in the forecasts, so the experiment conducted here investigates the impact of this on POD and FAR, as well as an increase in variance relative to the baseline case.

The upper and lower bounds for perturbations to the mean and variance of $\hat{y}$ are given in Table 5.10.

Bias in the mean

Figure 5.15 shows the results of the sensitivity experiment with the forecast mean for the DFDR (Figure 5.15a and b) and PTDR and PDR (Figure 5.15c and d) based on the FTC and FPC criteria. The immediate impression gained from Figure 5.15 is that there is strong sensitivity for the deterministic case (Figures 5.15a and b) while there is weak sensitivity for the probabilistic case (Figures 5.15c and d). The results are now analysed in this order.

a) Deterministic forecasts

FTC Criterion: A negative bias makes the values of $\hat{y}$ lower than the values of y. It reduces the number of hits and false alarms and increases the misses, and therefore reduces the POD and FAR values (Figure 5.15 and Table 5.11). Positive bias produces the contrary effect, i.e., it makes the values of $\hat{y}$ greater than the values of y and makes the POD and FAR values increase.

FPC Criterion: To recapitulate, in the FPC case, a hit represents a flooded house that was warned, a miss represents a flooded house that was not warned, and a false alarm represents a house that was warned and not flooded. The FPC results are different from the FTC results because FPC considers, apart from the misses and false alarms produced due to "wrong decisions" (misses and false alarms in the FTC), "additional" misses or false alarms existing after a "good" decision, i.e., when a warning in the at-risk community preceded a flood, but the event magnitude was incorrectly forecast (considered a hit in the FTC).

The results show that, for the FPC-based POD, the greater the negative bias of the mean, the lower the weight of misses due to "wrong decisions" (misses in the FTC)

in this metric. In other words, the negative impact of the "additional" misses considered in the FPC ***decreases*** as the negative bias of the mean increases. Figure 5.15a shows that there is a point when FPC-based POD equals FTC-based POD, which means that the negative impact of the "additional" misses on the FPC-based POD is negligible. The FPC-based FAR shows different behaviour. The negative impact of the "additional" false alarms considered in the FPC ***increases*** as the negative bias of the mean increases.

When there is a positive bias, the opposite effect occurs. The negative impact of the "additional" misses considered in the FPC ***increases*** as the positive bias of the mean increases. In addition, the negative impact of the "additional" false alarms considered in the FPC ***decreases*** as the positive bias of the mean increases. Figure 5.15a shows that there is a point when FPC-based FAR equals FTC-based FAR, which means that the negative impact of the "additional" false alarms on the FPC-based FAR is negligible.

a) Probabilistic Forecasts

FTC Criterion: As already observed, the POD and FAR results are largely insensitive to perturbations in the mean. To understand this result, it is necessary to consider what is happening to the bivariate distribution of $(y, \hat{y})$ and the predictive density $f(y|\hat{y})$. Figure 5.16 shows plots of the bivariate distributions corresponding to the lower bound (Figure 5.16a) and the upper bound (Figure 5.16b) perturbations. The baseline case is also shown in the plots. The predictive density for a given value of $\hat{y}$ is obtained by slicing the bivariate distribution in the vertical, as shown in Figure 5.8. The mean of the predictive density is $\hat{y}$, and once that is defined, the predictive density is controlled by the variability of y, but this, and the marginal distribution of y remain unchanged in this experiment. Figure 5.16 illustrates this graphically where the bivariate plots for the perturbations shift left or right for the negative and positive perturbations, respectively. This can be explained further by considering the expression for the mean of the predictive density in the Normal case, which is reproduced here from equation (Eq. 4.52) as

$$\mu_{x|\hat{x}} = \mu_z + \rho \frac{\sigma_z}{\sigma_{\hat{z}}} [ln(\hat{x}) - \mu_{\hat{z}}]$$

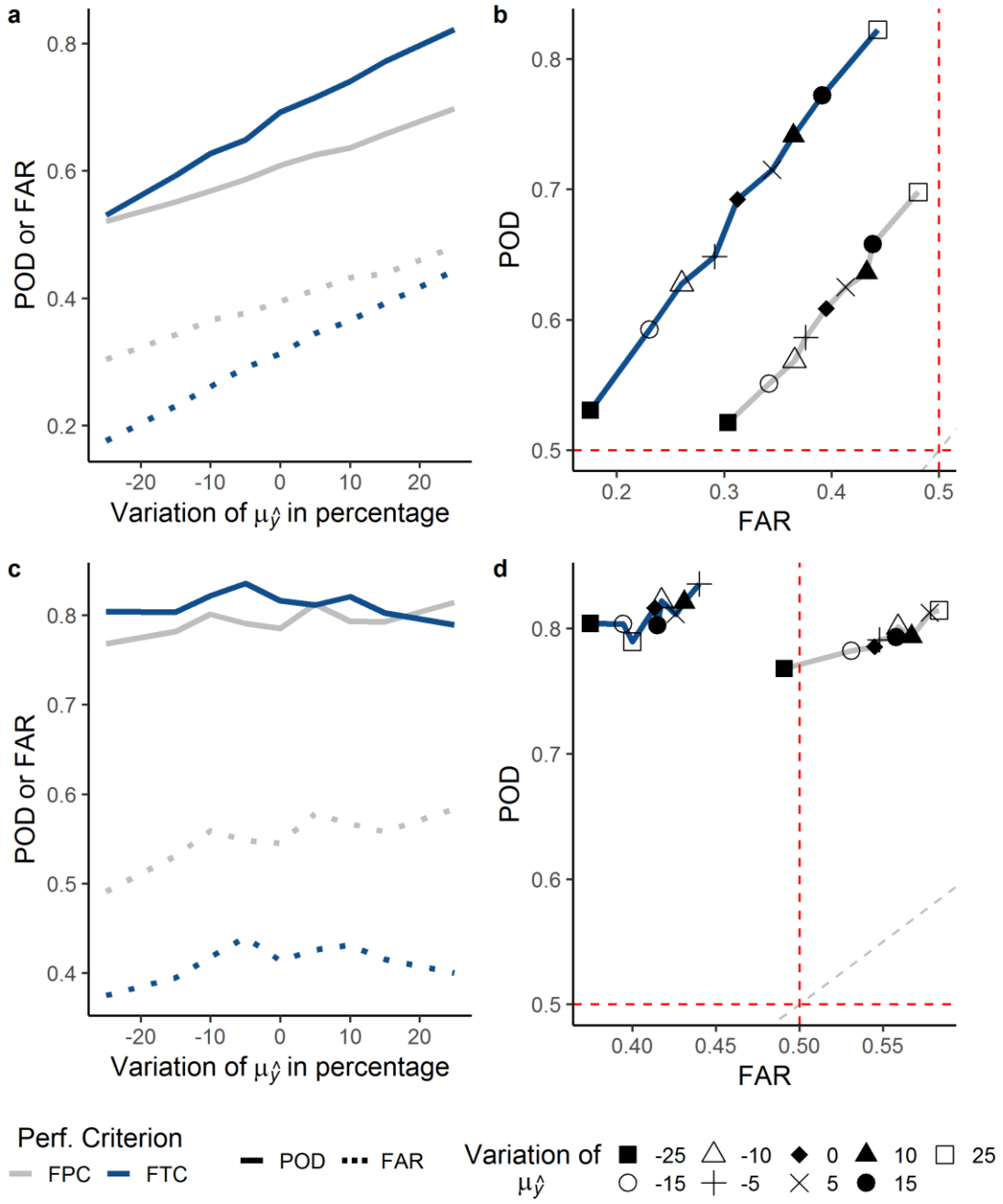


Figure 5.15: Results of the sensitivity experiment for $\mu_{\hat{y}}$ for the flooding threshold-based and floodplain property-based criterion, FTC and FPC, respectively, based on the deterministic rule DFDR, and probabilistic rules PTDR and PDR.

Figure a shows FAR and POD results for different variations of $\mu_{\hat{y}}$ based on the DFDR, whereas Figure c shows these values for the PTDR and PDR. FTC-based results (blue line) are shown for the DFDR and PTDR, and FPC-based results (grey line) are shown for the DFDR and PDR. The values of these figures are summarised in the FAR-POD curves shown in b and d, respectively.

POD: Probability of detection; FAR: false alarm ratio; PT: probabilistic threshold, DFDR: deterministic forecast-based decision rule, PTDR: probabilistic rule based on a PT and the probability of exceedance (PE) of y_T, PDR: probabilistic rule based on a PT and PE of y_w, FTC: flooding threshold-based criterion, FPC: floodplain property based criterion.

Table 5.11: Contingency table used to build the plots of POD and FAR in Figure 5.15

h, m, f, cn averaged over 30 replications.

Variation of $\mu_{\hat{y}}$	PT	Decision	FTC				FPC			FTC		FPC	
			h	m	f	cn	h	m	f	POD	FAR	POD	FAR
-0.25		DFDR	667.25	591.05	142.25	8599.5	146660	135444	64076	0.53	0.18	0.52	0.30
-0.15		DFDR	742.4	510.7	221.55	8525.4	154851	126050	80479	0.59	0.23	0.55	0.34
-0.1		DFDR	782.15	464.6	275.55	8477.7	158075	119941	91266	0.63	0.26	0.57	0.37
-0.05		DFDR	817.3	441.5	334.2	8407	167184	117329	101036	0.65	0.29	0.59	0.38
0		DFDR	869.35	386.05	395.8	8348.8	171716	110644	112030	0.70	0.31	0.61	0.40
0.05		DFDR	886.2	353.1	465.4	8295.3	173886	104115	122879	0.72	0.34	0.62	0.41
0.1		DFDR	927.6	323.6	531.35	8217.5	176779	101010	134280	0.74	0.36	0.64	0.43
0.15		DFDR	966.3	283.7	622.6	8127.4	185704	96093	144924	0.77	0.39	0.66	0.44
0.25		DFDR	1029.7	223.75	821.9	7924.7	195418	84666	180995	0.82	0.44	0.70	0.48
-0.25	0.63	PTDR	1012.9	245.45	623.7	8118				0.80	0.38		
-0.15	0.63	PTDR	1007.7	245.4	674.05	8072.9				0.80	0.39		
-0.1	0.62	PTDR	1024.4	222.35	762.4	7990.9				0.82	0.42		
-0.05	0.61	PTDR	1052.7	206.1	845.95	7895.3				0.84	0.44		
0	0.63	PTDR	1024.4	231.05	751.4	7993.2				0.82	0.41		
0.05	0.63	PTDR	1004.9	234.4	770.85	7989.9				0.82	0.43		
0.1	0.63	PTDR	1027.1	224.15	801.75	7947.1				0.82	0.43		
0.15	0.63	PTDR	1003.8	246.2	747.75	8002.3				0.80	0.42		
0.25	0.64	PTDR	989.65	263.75	691.45	8055.2				0.79	0.40		
-0.25	0.60	PDR	1076.2	182.1	856.95	7884.8	216502	65603	219893			0.77	0.49
-0.15	0.59	PDR	1088.5	164.65	992.35	7754.6	219574	61327	257640			0.78	0.53
-0.1	0.58	PDR	1105.8	140.95	1098.9	7654.4	222747	55269	291396			0.80	0.56
-0.05	0.58	PDR	1104.3	154.5	1107	7634.2	224665	59848	291191			0.79	0.55
0	0.59	PDR	1100.2	155.2	1050.4	7694.2	221480	60880	273767			0.79	0.55
0.05	0.57	PDR	1100.7	138.6	1218.1	7542.6	225803	52199	323607			0.81	0.58
0.1	0.58	PDR	1103.8	147.4	1139.4	7609.5	220750	57039	302536			0.79	0.57
0.15	0.58	PDR	1099.4	150.65	1140.3	7609.7	223162	58634	294506			0.79	0.56
0.25	0.57	PDR	1116.5	136.9	1237.4	7509.2	228067	52017	336869			0.81	0.58

So for a given $\hat{x}$ (recall that $\hat{x}$ = ($\hat{y}$ - $\hat{y}_o$) and $\hat{z} = ln(\hat{x})$, the mean of the predictive density will be shifted up or down relative to μ_z by the second term in brackets on the rhs, scaled by $\rho \frac{\sigma_z}{\sigma_{\hat{z}}}$ which is a constant in this experiment. The mean of this term will be zero across all the forecasts considered, which means that the optimal probabilistic threshold PT and POD and FAR will not change as $\mu_{\hat{z}}$ and $\mu_{\hat{y}}$ change, bearing in mind that the conditional mean in the Normal space is related to the mean in the real space (Table 4.1) as:

$$\mu_{y|\hat{y}} = \hat{y}_o + exp(\mu_{z|\hat{z}} + 0.5\sigma^2_{z|\hat{z}})$$

The minor fluctuations shown in Figure 5.15 are due to sampling variability.

As a consequence of this lack of sensitivity, the optimal POD and FAR values for all the perturbations cluster closely around the baseline case in the POD-FAR plot, with all POD values around 0.8 and FAR values around 0.4. This is in contrast to the deterministic forecast case where POD degrades towards 0.5 for the lower bound, but on the other hand, POD increases to just above 0.8 at the upper bound, with FAR tending towards 0.5. Therefore, probabilistic forecasting can be seen to be robust to biases in the mean by maintaining stability in POD and FAR, while this is not the case for deterministic forecasting. However, the results obtained here suggest that there might be merit in increasing the mean of $\hat{y}$ to be greater than that of y in deterministic forecasting models to get higher values of POD.

FPC Criterion: the results for this criterion mimic those for the FTC case in terms of lack of sensitivity, with POD being slightly lower than for the FTC case, but with much higher values of FAR, thus placing nearly all the values of POD and FAR in a cluster in the upper right quadrant (POD>0.5,FAR>0.5). Therefore, probabilistic forecasting is also robust to biases in the mean of $\hat{y}$ based on the FPC criterion, while deterministic forecasting is not. However, the deterministic results suggest that increasing the mean of $\hat{y}$ relative to that of y can lead to higher values of POD, but with increasing FAR.

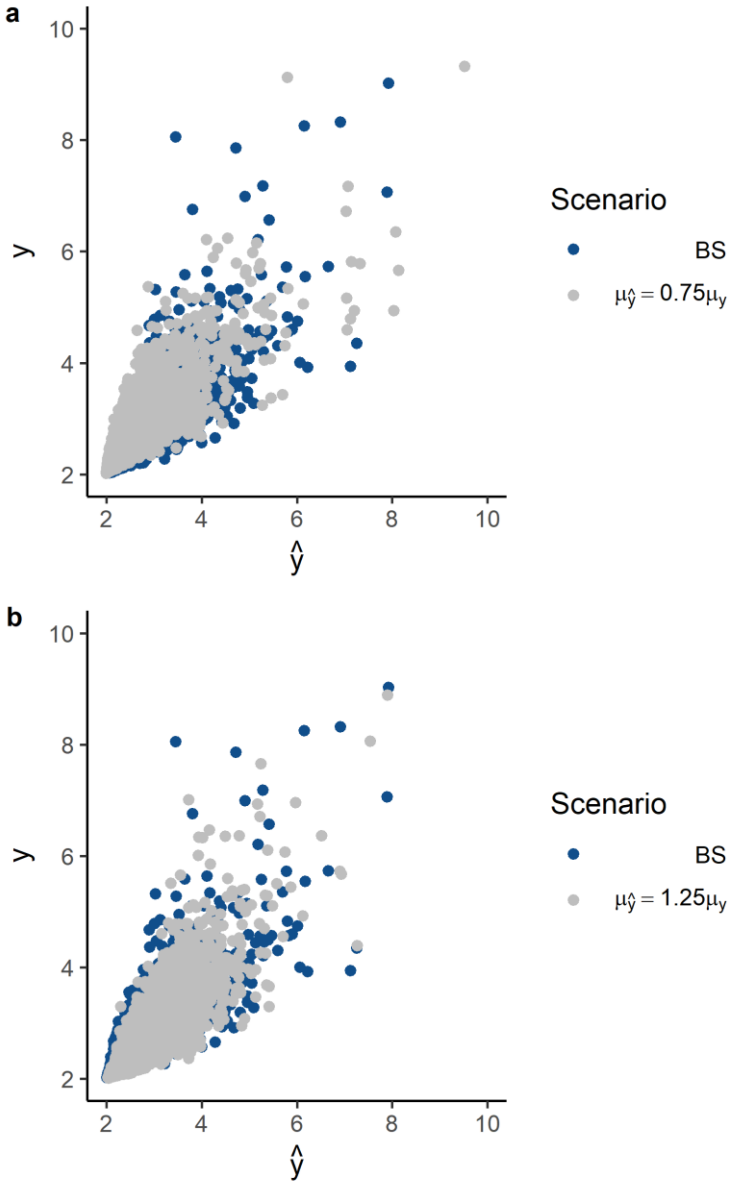


Figure 5.16: Pairs $(y,\hat{y})$ for the baseline scenario and for the cases of the lower (a) and upper (b) bounds of the sensitivity experiment of $\mu_{\hat{y}}$

These bivariate plots represent one replication of 10,000 values of the thirty used to build Figure 5.15

BS: Baseline scenario.

Bias in the variance

Figure 5.17 shows the results of the sensitivity experiment with the forecast variance for the DFDR (Figure 5.17a and Figure 5.17b) and PTDR and PDR (Figure 5.17c and Figure 5.17d) based on the FTC and FPC criteria. The values of this Figure are described in Table 5.12. The immediate impression gained from Figure 5.17 is that there is strong sensitivity for the deterministic case (Figure 5.17 and b) while there is very weak sensitivity for the probabilistic case (Figure 5.17c and d). The results are now analysed in this order.

a) **Deterministic forecasts**

FTC and FPC Criteria: The results in Figure 5.17a show that, for the FPC-based POD, the greater the negative bias in the variance, the greater the weight of misses after "good" decisions (hits in the FTC) on this metric. In other words, the negative impact of "additional misses" considered in the FPC case increases as the negative bias in the variance increases (Figure 5.17a). The FPC-based FAR, however, shows different behaviour, i.e., the negative impact of "additional false alarms" considered in the FPC ***decreases*** as the negative bias in the variance increases (Figure 5.17a). Figure 5.17a also shows that there is a point when FPC-based FAR equals FTC-based FAR, which means that the negative impact of the "additional" false alarms on the FPC-based FAR is negligible.

When there is a positive bias in the variance, the opposite effect occurs. The negative impact of the "additional" misses considered in the FPC ***decreases*** as the positive bias increases. Figure 5.17a shows that there is a point when FPC-based POD equals FTC-based POD, which means that the negative impact of the "additional" misses on the FPC-based POD is negligible. The FPC-based FAR shows different behaviour. The negative impact of "additional false alarms" considered in the FPC ***increases*** as the positive bias in the variance increases (Figure 5.17a).

The POD-FAR plot in Figure 5.17b shows again that the FPC plot is below that of the FTC, reflecting the more demanding FPC criterion. In the FPC case, POD and FAR show a much wider range of variation than FTC, with a negative bias in the variance resulting in values in the lower left quadrant. On the other hand, a positive bias increases the POD value towards 0.7, but the FAR value also increases towards 0.5.

A comparison of Figure 5.17b with Figure 5.15b shows that in the FTC case, there is higher sensitivity to a bias in the mean than for FPC, while the opposite is the case for bias in the variance.

a) **Probabilistic Forecasts**

FTC Criterion: As already observed, the POD and FAR results are largely insensitive to perturbations in the variance for both the FTC and FPC criteria. To

understand this result, it is necessary, as in the case of the mean, to consider what is happening to the bivariate distribution of $(y, \hat{y})$ and the predictive density $f(y|\hat{y})$. Figure 5.18 shows plots of the bivariate distributions corresponding to the lower bound (Figure 5.18a) and the upper bound (Figure 5.18b) perturbations. The baseline case is also shown in the plots. The predictive density for a given value of $\hat{y}$ is obtained by slicing the bivariate distribution in the vertical, as shown in Figure 5.8. The mean of the predictive density is $\hat{y}$, and once that is defined, the predictive density is largely controlled by the variability of y, but this, and the marginal distribution of y remain unchanged in this experiment. Figure 5.18 illustrates this graphically where the bivariate plots for the perturbations contract and expand along the x-axis for the negative and positive perturbations, respectively. This can be explained further by considering the expression for the variance of the predictive density in the Normal space which is reproduced here from equation (4.53) as

$$\sigma_{x|\hat{x}} = \sigma_z\sqrt{1-\rho^2}$$

So the variance of the predictive density in the Normal space is uniquely controlled by the variance of z, as the correlation ρ is unchanged in this experiment. This means that the optimal probabilistic threshold PT and POD and FAR will not change as $\sigma_{\hat{z}}^2$ and $\sigma_{\hat{y}}^2$ change, bearing in mind that the conditional variance in the Normal space is related to that in the real space (Table 4.1) as:

$$\sigma_{y|\hat{y}}^2 = exp\left(\sigma_{z|\hat{z}}^2 + 2\mu_{z|\hat{z}}\right)\left[exp\left(\sigma_{z|\hat{z}}^2\right) - 1\right],$$

and that $\mu_{\hat{z}}$ is unchanged for this experiment.

The minor fluctuations shown in Figure 5.17 are due to sampling variability.

As a consequence of this lack of sensitivity, the optimal POD and FAR values for all the perturbations cluster closely around the baseline case in the POD-FAR plot, with all POD values just above 0.8 and FAR values around 0.4. This is in contrast to the deterministic forecast case where POD degrades towards 0.6 for the lower bound and to just above 0.7 at the upper bound, with FAR tending towards 0.3. Therefore, as in the case of the mean, probabilistic forecasting can be seen to be robust to bias in the variance by maintaining stability in POD and FAR, while this is not the case for deterministic forecasting. However, the results obtained here suggest that there might be some merit in increasing the variance of $\hat{y}$ to be equal

to or greater than that of y in deterministic forecasting models to get higher values of POD.

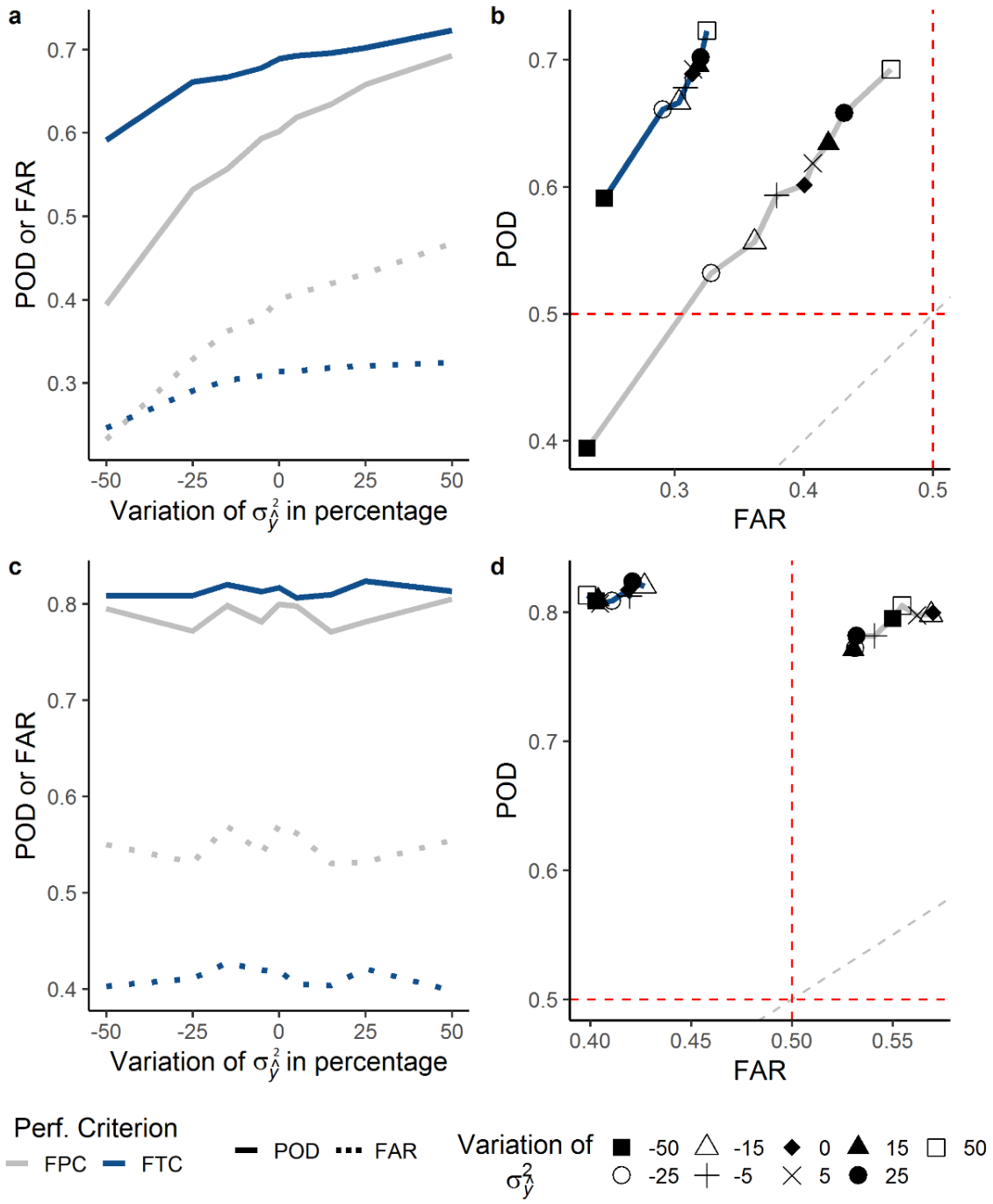


Figure 5.17: Results of the sensitivity experiment for $\sigma^2_{\hat{y}}$ for the flooding threshold-based and floodplain property-based criterion, FTC and FPC, respectively, based on the deterministic rule DFDR, and probabilistic rules PTDR and PDR.

Figures a shows POD and FAR results for different variations of $\sigma^2_{\hat{y}}$ based on the DFDR, whereas Figure c shows these values for the PTDR and PDR. FTC-based results (blue line) are shown for the DFDR and PTDR, and FPC-based results (grey line) are shown for the DFDR and PDR. The values of these figures are summarised in the POD-FAR curves shown in b and d, respectively.

POD: Probability of detection; FAR: false alarm ratio; PT: probabilistic threshold, DFDR: deterministic forecast-based decision rule, PTDR: probabilistic rule based on a PT and the probability of exceedance (PE) of y_T, PDR: probabilistic rule based on a PT and PE of y_w, FTC: flooding threshold-based criterion, FPC: floodplain property based criterion.

Table 5.12: Contingency table used to build the plots of POD and FAR in Figure 5.17

h, m, f, cn averaged over 30 replications.

Variation of	PT	Decision	FTC				FPC			FTC		FPC	
			H	m	f	Cn	h	m	f	POD	FAR	POD	FAR
-0.5		DFDR	748.85	516.85	243.8	8490.5	111581	171468	33933	0.59	0.246	0.39	0.23
-0.25		DFDR	827.2	423.95	340.45	8408.4	151198	132433	74061	0.66	0.291	0.53	0.33
-0.15		DFDR	834.2	417.5	364.3	8384	155661	124191	88184	0.67	0.3035	0.56	0.36
-0.05		DFDR	849.4	402.35	377.7	8370.6	169254	115800	103354	0.68	0.3085	0.59	0.38
0		DFDR	860.75	388.3	393.15	8357.8	166485	110574	111090	0.69	0.314	0.60	0.40
0.05		DFDR	868.55	387.4	400.7	8343.4	174652	107950	119561	0.69	0.3145	0.62	0.41
0.15		DFDR	872.65	379.7	406.4	8341.3	180153	103605	129422	0.69	0.3185	0.63	0.42
0.25		DFDR	886.35	376.35	419.4	8317.9	187547	96908	142766	0.70	0.3205	0.66	0.43
0.5		DFDR	900.85	346.4	434.5	8318.3	192498	85472	169217	0.72	0.325	0.69	0.47
-0.5	0.63	PTDR	1022.6	243.1	722.9	8011.4				0.81	0.40		
-0.25	0.63	PTDR	1011.6	239.6	735.65	8013.2				0.81	0.41		
-0.15	0.63	PTDR	1027	224.75	791.25	7957.1				0.82	0.43		
-0.05	0.63	PTDR	1016.4	235.35	770.15	7978.1				0.81	0.42		
0	0.63	PTDR	1021.2	227.9	751.4	7999.6				0.82	0.42		
0.05	0.63	PTDR	1012.5	243.5	715.05	8029				0.81	0.41		
0.15	0.63	PTDR	1013.7	238.7	709.5	8038.2				0.81	0.40		
0.25	0.62	PTDR	1040.2	222.5	775.85	7961.5				0.82	0.42		
0.5	0.63	PTDR	1014.5	232.8	684.75	8068				0.81	0.40		
-0.5	0.58	PDR	1116.5	149.25	1098.9	7635.4	225300	57748	288296			0.80	0.55
-0.25	0.60	PDR	1081.3	169.85	1004.7	7744.2	218747	64885	258115			0.78	0.53
-0.15	0.58	PDR	1108.6	143.1	1144.9	7603.4	223263	56590	299956			0.80	0.57
-0.05	0.59	PDR	1089.2	162.55	1065.6	7682.7	222940	62114	276753			0.78	0.54
0	0.58	PDR	1109.2	139.85	1150.7	7600.3	222143	54916	306534			0.80	0.57
0.05	0.58	PDR	1112.3	143.65	1138.7	7605.4	225486	57115	302658			0.80	0.56
0.15	0.59	PDR	1087.1	165.3	1040.5	7707.2	218713	65045	268172			0.78	0.53
0.25	0.59	PDR	1097.7	165.05	1018	7719.4	222144	62310	262491			0.78	0.53
0.5	0.58	PDR	1109	138.25	1086.3	7666.5	223423	54547	285619			0.80	0.55

FPC Criterion: The results for this criterion mimic those for the FTC case in terms of lack of sensitivity, with POD being slightly lower than for the FTC case, but with much higher values of FAR, thus placing nearly all the values of POD and FAR in a cluster in the upper right quadrant (POD>0.5,FAR>0.5). Therefore, probabilistic forecasting is also robust to a bias in the variance of $\hat{y}$ based on the FPC criterion, while deterministic forecasting is not. In particular, the POD values degrade substantially towards the lower bound, suggesting that increasing the variance of $\hat{y}$ to be equal to or greater than that of y can lead to higher values of POD, but with increasing FAR.

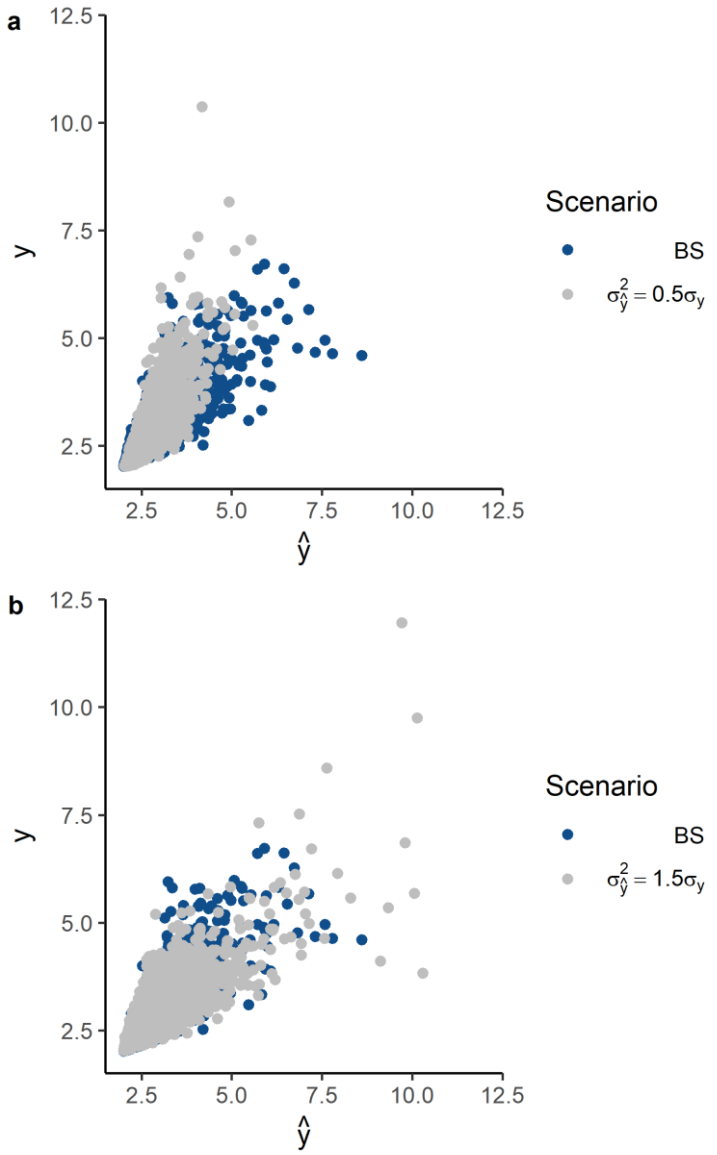


Figure 5.18: Pairs $(y,\hat{y})$ for the baseline scenario and for the cases of the lower (a) and upper (b) bounds of the sensitivity experiment of $\sigma^2_{\hat{y}}$

These bivariate plots represent one replication of 10,000 values of the thirty used to build Figure 5.17

BS: Baseline scenario.

5.7.2.4 Relevance to the case of Morpeth flooding

The baseline scenario is based on the statistics of observed peak water levels derived for the Mitford gauging station on the River Wansbeck which is adjacent to Morpeth (Section 4.5.2). However, it was intended that the baseline should be generic and not be viewed as particular to a FEWS system for any specific location. However, by modifying the SoP from a 5-year to a 10-year flood, this reflects the Morpeth case more accurately and allows an assessment of what level of correlation between forecast and observed peak levels might be needed to provide a FEWS with a medium to high warning standard, as defined in section 5.7.2.1 above for POD and FAR. Moreover, it allows the effect of increasing the SoP on POD and FAR to be assessed.

Figure 5.19 shows the effect of an increase in the SoP from 5 to 10 years on the behaviour of POD and FAR; this figure is directly comparable to Figure 5.9 for the baseline case. As the correlation between the forecast and observed values will decrease as a higher threshold (SoP) is applied, the overall levels of POD drop for the FTC criterion, but interestingly, the POD-FAR curve in Figure 5.19c for the FPC criterion is closer to that for the FTC criterion than for the baseline case (Figure 5.9c). This suggests that the results for FPC are less affected than FTC for the SoP of 10 years.

Table 5.13: Optimal values of POD and FAR for the three cases described by Figures 5.9, 5.19 and 5.20

Case	Prob. threshold	POD	FAR	POD- FAR	Decision rule	Criterion
Baseline (Figure 5.9)	-	0.68	0.32	0.36	DFDR	FTC
	-	0.60	0.40	0.20		FPC
	0.62	0.83	0.44	0.39	PTDR	FTC
	0.58	0.80	0.57	0.23	PDR	FPC
Morpeth case (SoP=10 years, $IP_\rho = 0.85$) (Figure 5.19)	-	0.63	0.37	0.26	DFDR	FTC
	-	0.58	0.42	0.16		FPC
	0.61	0.78	0.48	0.30	PTDR	FTC
	0.55	0.82	0.63	0.19	PDR	FPC
Morpeth case (SoP=10 years, $IP_\rho = 0.94$) (Figure 5.20)	-	0.76	0.24	0.52	DFDR	FTC
	-	0.72	0.27	0.45	DFDR	FPC
	0.60	0.80	0.27	0.53	PTDR	FTC
	0.58	0.77	0.32	0.45	PDR	FPC

Table 5.13 gives the optimal values of POD and FAR that are shown graphically in Figures 5.19a and 5.19b; for probabilistic forecasting, they are POD = 0.78 and FAR 0.48 (FTC) and POD = 0.82 and FAR = 0.63 (FPC). The values of FAR are too high to be classified in the Medium Performance category, so a further experiment was performed in which the IP_ρ value (correlation for a lead time of 6 hours) was increased by 10% to 0.94. The corresponding metrics are POD = 0.80 and FAR 0.27 (FTC) and POD = 0.77 and FAR = 0.32 (FPC), showing significant reductions in FAR, and placing the performance in the Medium to High category. This would require a model with a high level of performance in forecasting peak water levels.

Overall, improving by 10% the assumed forecast performance of Morpeth case (expressed as IP_ρ=0.94 and no biases in the mean and variance), the reliability of flood warning improves significantly under both the deterministic and probabilistic scenarios (Table 5.13). The results of experiment 2 show that this significant improvement occurs only in FEWSs with relatively good forecast performance; the assumed forecast performance of the Morpeth case falls in this category.

The advantage of using probabilistic forecasting is more noticeable in FEWS with considerable forecast uncertainty. If one assumes that an optimal FEWS is obtained by optimizing the difference between POD and FAR, Table 5.13 shows that when the forecast performance of a FEWS is very good (expressed as IP_ρ=0.94 and no biases in the mean and variance), the reliability of flood warnings under the deterministic and probabilistic scenario is practically the same.

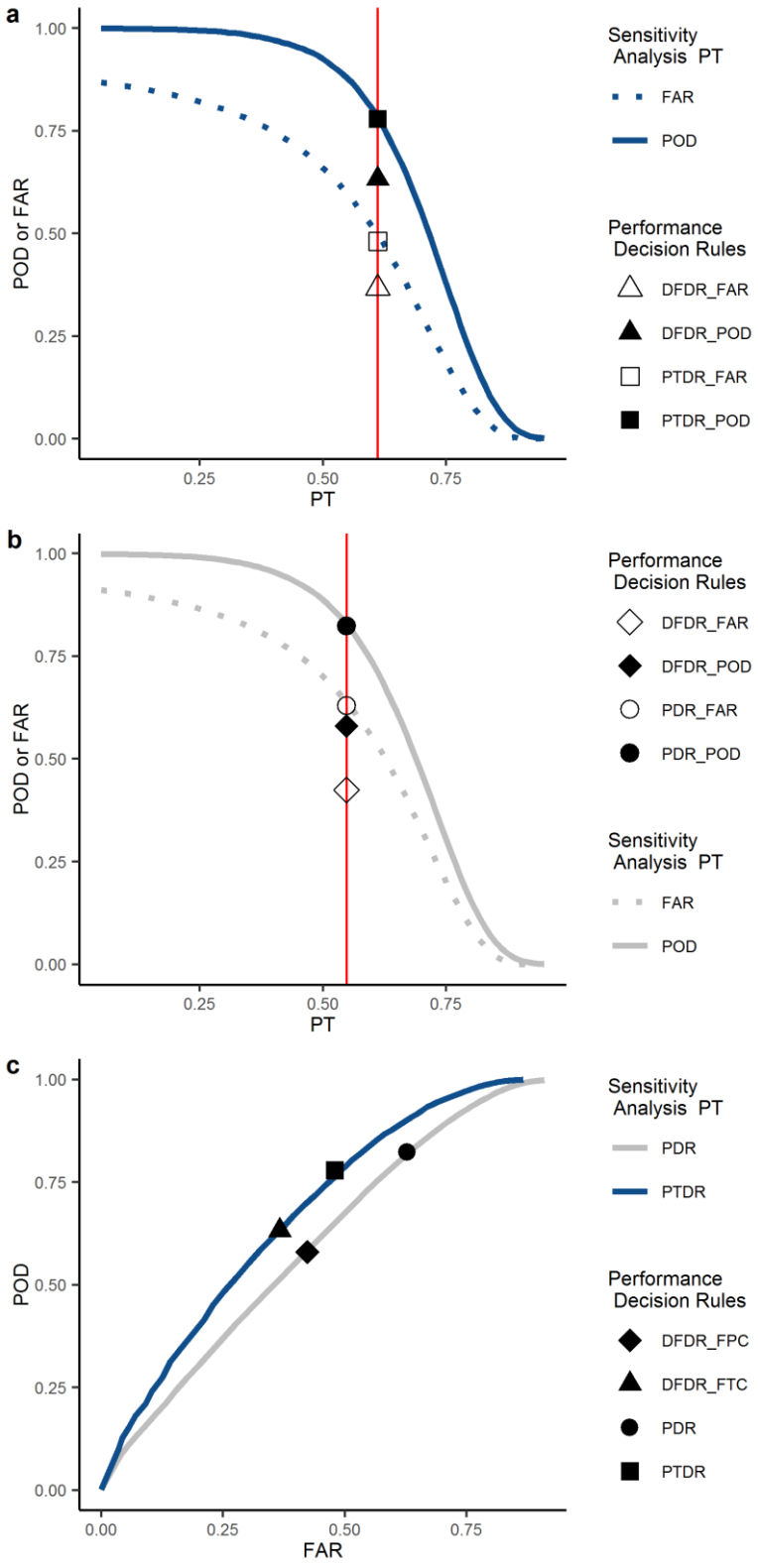


Figure 5.19: Analysis of the performance of the baseline FEWS for the flooding threshold-based and floodplain property-based criterion, FTC and FPC, respectively, by assuming a standard of protection SoP of 10 years

This figure shows the results one would obtain by perturbing the baseline case (Figure 5.9) by assuming an SoP of 10 years.

POD: Probability of detection; FAR: false alarm ratio; PT: probabilistic threshold, DFDR: deterministic forecast-based decision rule (DFDR), PTDR: probabilistic rule based on a PT and the probability of exceedance (PE) of y_T, PDR: probabilistic rule based on a PT and the PE of y_w.

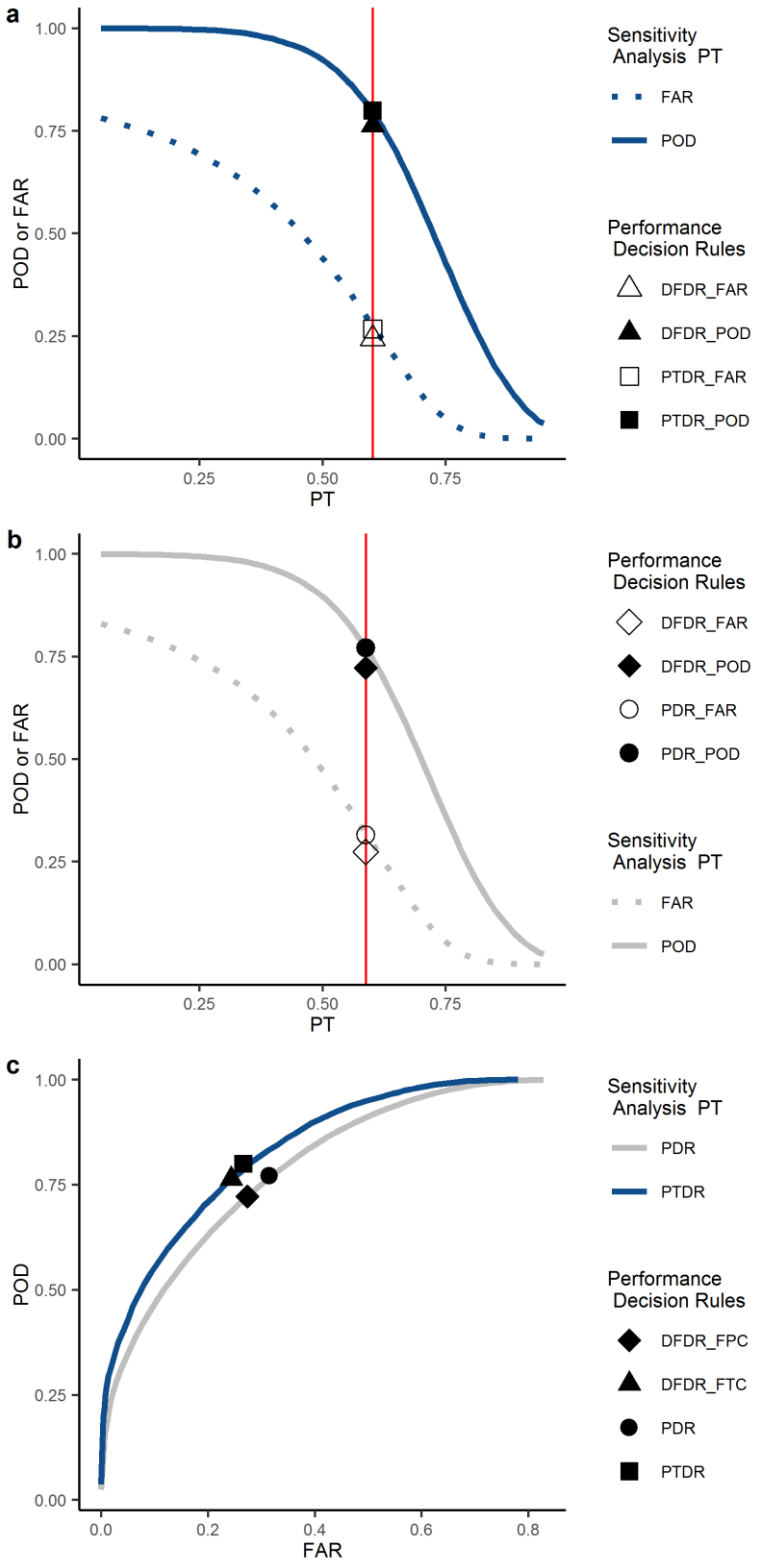


Figure 5.20: Analysis of the performance of the baseline FEWS for the flooding threshold-based and floodplain-property based, FTC and FPC, respectively, by assuming a standard of protection SoP of 10 years and IP_ρ=0.94 (correlation value to the catchment lag L).

This figure shows the results one would obtain by perturbing the baseline case (Figure 5.9) by assuming an SoP of 10 years and IP_ρ=0.94.

POD: Probability of detection; FAR: false alarm ratio; PT: probabilistic threshold, DFDR: deterministic forecast-based decision rule (DFDR), PTDR: probabilistic rule based on a PT and the probability of exceedance (PE) of y_T, PDR: probabilistic rule based on a PT and the PE of y_w.

5.8 Main Findings

A threshold-based approach has been used as a warning criterion, where deterministic and probabilistic decision rules were used to simulate them. The probabilistic-based decision rule PDR used to simulate FEWSs with real-time flood maps is novel. This rule uses a warning level $\hat{y}_w$ derived from $f(y|\hat{y})$ to make the warning decision, where $\hat{y}_w$ is defined by a probabilistic threshold PT, which has to be optimised based on a pre-defined criterion. The PDR was used because FEWS based on real-time flood maps must use a magnitude's prediction of the potential flood to generate them. Impact curves defining the number of affected houses can be used for exploring the flood warning's reliability in terms of affected houses (the FPC). An impact curve can be built with few flood events, and its use exploits the process done by inundation models well and contributes to the integrated framework's versatility defining a FEWS.

It was analysed that the performance based on the FPC is lower than that obtained based on the FTC. That difference was directly related to the uncertainty of the flood magnitude, i.e., the difference between y and $\hat{y}$, which defines the difference between the warned and flooded properties. In this context, it was concluded that the flood magnitude's uncertainty is an important factor influencing the flood warning reliability of a FEWS.

The warning strategy is another important factor influencing the reliability of flood warnings. This research showed that a deterministic-based warning strategy in the FEWS produces sub-optimal decisions and that a probabilistic-based warning strategy, where the forecast errors are acknowledged, can use an optimization criterion to improve the reliability of flood warnings. In this research, the biggest difference between POD and FAR was used as an optimization criterion. The results showed that an optimal warning strategy based on this criterion tends to deliver FAR values greater than the one one would obtain based on deterministic forecasts. Thus, this strategy can be used when one wants to give more weight to the reduction of missed events whose economic consequences, in terms of floods, are far greater than those associated with false alarms. The forecasting lead time τ was also identified as important factor influencing the flood warning reliability. As expected, flood warning reliability declines with lead time according to the performance function used, which shows a faster decline in correlation for lead times greater

than the catchment lag, reflecting the greater uncertainty resulting from QPFs. In particular, probabilistic forecasting copes much better with the increasing uncertainty than deterministic forecasting, where the POD values are much higher for the probabilistic case but at the expense of higher FAR values. The SoP was a factor impacting the reliability of the FEWS. As the correlation between the forecast and observed values will decrease as a higher threshold SoP is applied, this flood warning attribute decreases as the SoP increases.

The biases in the forecast mean and variance were also identified as important factors influencing the flood warning reliability. Remarkable robustness to biases in these two variates has been observed for probabilistic forecasting, which is a consequence of using PU, which is based on the conditional density of y given $\hat{y}$, whereas deterministic forecasting shows high sensitivity. The results for the latter case suggest that increasing the mean and variance of the forecasts relative to those of the observed could improve reliability in this case (Figures 5.17a and b).

The correlation between the observed and forecast peak water levels has been shown to be an important factor controlling flood warning reliability. The analysis done in this Chapter (section 5.7.2.2) suggests that if the forecast performance is mainly controlled by $\rho_{y\hat{y}}$ (a bias in the mean or variance of the forecasts can also affect performance), an improvement of 10% of this factor, in terms of flood warning reliability, is more beneficial in FEWSs with relatively small than considerable forecast uncertainty. Furthermore, some indicative results for the case of a hypothetical Morpeth FEWS suggest that a correlation of 0.94 in peak discharges would be needed to obtain POD and FAR values in the Medium to High category for probabilistic forecasting.

Chapter 6. Hydro-economic modelling of the benefits of flood early warning systems (FEWS)

6.1 Introduction

In chapter 5, the generic framework was partially used to explore the performance of a river flood early warning system (FEWS) in terms of reliability. The Monte Carlo flood and forecast generator (MCFG), the flood warning-decision making model (FWDC), and part of the response and impact component (RIC) were considered to analyse this flood warning performance attribute based on two performance criteria and considering deterministic- and probabilistic-based warning decisions.

In this Chapter, the generic framework is employed fully to build a hydro-economic expected damage (ED) model (Figure 6.1a) which is used to explore the performance of an inundation forecasting-based FEWS in terms of both economic effectiveness and reliability. The hydro-economic ED model is based on three components: i) the MCFG that defines the univariate and bivariate probability of the magnitude of flood threats, i.e., peak water levels (y) and their forecasts ($\hat{y}$), ii) the FWDC that simulates warning decisions in the FEWS, and iii) the RIC that simulates the 'actual' and forecast flood impact in the floodplain, considering the flood extent and depth as the main variates measuring damage. These three linked components are used to do a Monte Carlo (MC) flood risk analysis in the no warning scenario (NWS), the perfect forecast scenario (PFS), and in the imperfect forecast scenarios, i.e., the deterministic forecast scenario (DFS) and the probabilistic forecast scenario (PrFS). This flood risk analysis provides a single aggregated value of the relationship between flood threats in the at-risk community and its economic consequences. That single value is the expected damage ED. In the NWS and PFS, part of the MCFG and RIC is used, and the ED is estimated from events generated from the probability distribution of y (Figure 6.1b and Figure 6.1c), whereas in the imperfect forecast scenarios (the DFS and PrFS), the three components of the hydro-economic ED model are fully exploited, and the ED is estimated from bivariate events generated from the joint distribution of y and $\hat{y}$ (Figure 6.1d). The ED in each scenario is obtained based on damage functions and the frequencies of that damage obtained from the MC results. Since the ED associated with the with-warning scenarios includes the cost of the warning

response (C_w), the word 'damage' in these scenarios is defined as the net damage that considers flood damage, mitigated and not mitigated by a proactive action, *and* C_w. The net damage associated with forecast uncertainty due to 'wrong' warning decisions (false alarms and misses) and 'good' warning decisions (hits) are considered in the analysis (this net damage is present in hits due to the difference between the warned and flooded houses).

Since the hydro-economic ED model in the imperfect forecast scenarios simulates, the warning decision and the resulting warned and flooded houses, it evaluates the reliability of flood warnings in terms of the floodplain property-based criterion (FPC) and the economic consequences. The reliability is evaluated in terms of the probability of detection (POD) and false alarm ratio (FAR).

The ED values are used to estimate the economic effectiveness of a FEWS as:

$$E_w = \frac{ED_{nw} - ED_w}{ED_{nw}} 100 \qquad \text{Eq. 6.1}$$

where E_w is the economic effectiveness of a FEWS, ED_{nw} is the ED of the at-risk community under the NWS, and ED_w is the ED of the at-risk community with the warning service. Since the ED is the metric used in this work to define the economic flood risk, The economic effectiveness E_w, in essence, represents the expected economic benefits associated with a warning service relative to the NWS.

The hydro-economic ED model is, thus, used to gain an insight into the relationship between the reliability (POD/FAR) and E_w of a FEWS , as these are different measures of performance, and into the impact of the forecast uncertainty on its economic benefits. That is done by evaluating the sensitivity of E_w to the main input parameters and controlling factors through a one-at-time method.

This chapter is structured as follows. In section 6.2, the concept of net damage associated with forecast uncertainty is explained. Then, from sections 6.3 to 6.6, the hydro-economic ED model is introduced. Thus, in section 6.3, the metrics and concepts used by the model to define the reliability and E_w of flood warnings are explained. In section 6.4, each component of the hydro-economic ED model is described. In section 6.5, the concept of flood risk and the basis of the MC flood risk analysis is introduced. In section 6.6, the algorithms of the hydro-economic ED

model to estimate the economic effectiveness E_w and reliability in terms of POD and FAR, are introduced. Finally, in section 6.7 the results of the sensitivity experiments are presented.

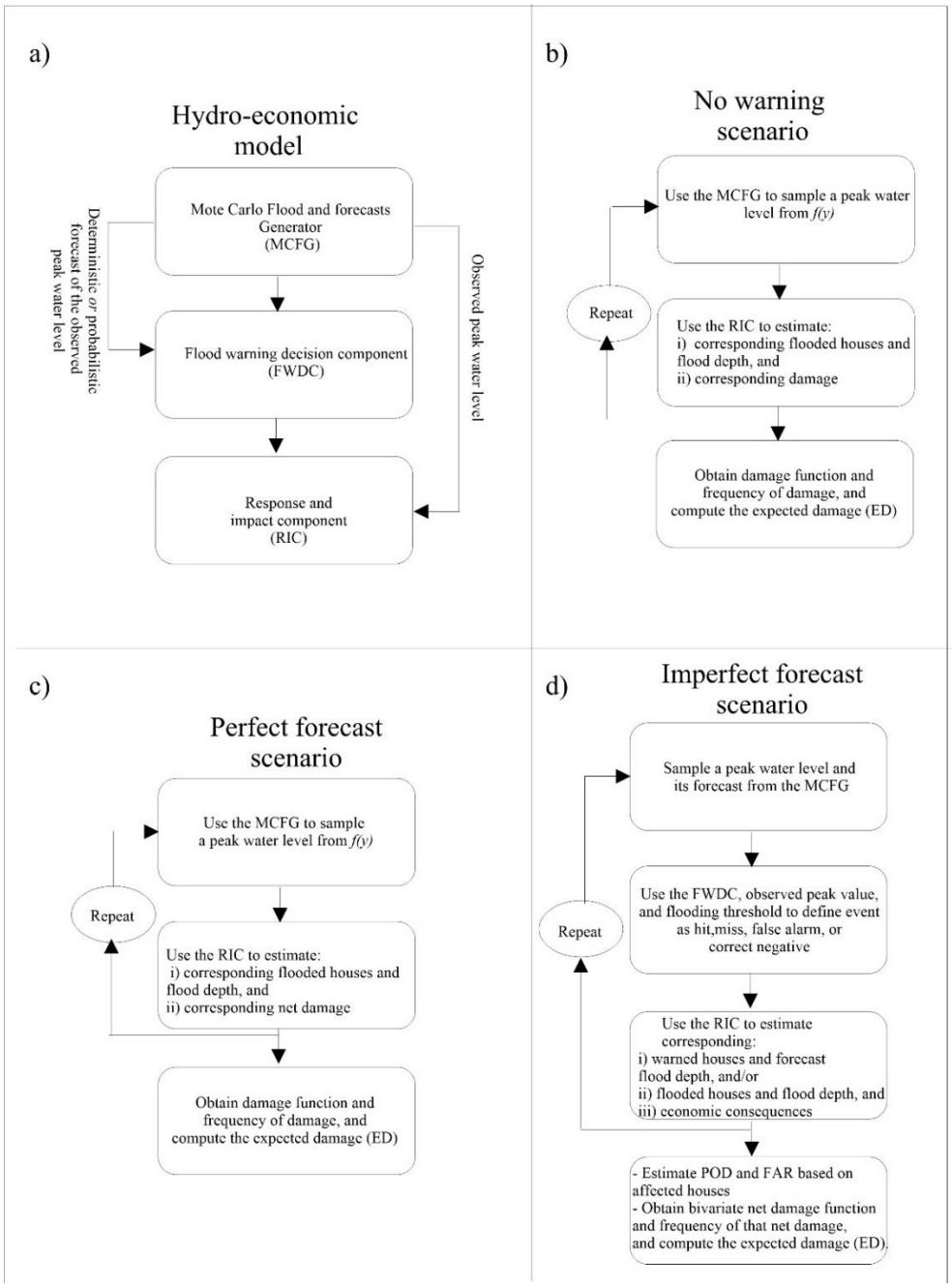


Figure 6.1: Illustration of the MC framework to explore the reliability and economic effectiveness of flood warnings.

The hydro-economic ED model has three main components (a) that are fully used only in the imperfect forecast scenarios (d). In each scenario, the model estimates the economic consequences within the full range of potential flood events in the at-risk community. The concept of expected damage (ED) is then used to obtain a single aggregated value of these consequences in the no warning scenario (b), perfect forecast scenario (c), and in an imperfect forecast scenario (d). The ED values are then used to estimate the flood warning's economic effectiveness as the ED reduction of a forecasting scenario relative to the NWS. The flood warning reliability is estimated in terms of POD and FAR based on warned and flooded houses

6.2 Net damage associated with forecast uncertainty

As was explained in the introductory part of this Chapter, the economic effectiveness E_w is evaluated in this work in terms of the expected economic benefits of a FEWS, and it should, therefore, include the costs incurred by the system. These costs should be somehow derived by considering the costs for setting up, operating, and maintaining the system, and the cost of the warning response C_w (per-event costs). All these costs, except C_w, can be considered 'fixed', and they can be included in the analysis of the expected economic benefits by prorating them through the economic lifespan of the system. The cost of the warning response C_w is, on the other hand, a cost incurred every time a warning is issued and depends on the magnitude of the forecast flooding and the lead time τ. It considers the cost of issuing a warning and any cost incurred in the dissemination process and by the public taking action due to the flood warning (proactive action). This cost is affected by the forecast uncertainty, and since this research is related to this topic, it is the only cost included in the analysis of E_w.

In an inundation-forecasting based FEWS, the warning decision is first made, and then, the warning dissemination is done based on the forecast of the flood extent. In this process, the number of houses warned (n_{houses}^{warned}) can be overestimated or underestimated due to the forecast uncertainty. They are overestimated or underestimated in the sense that the system warns *more or fewer* houses than it should warn, i.e., n_{houses}^{warned} is *higher or lower* than the number of houses flooded ($n_{houses}^{flooded}$), respectively. A FEWS mitigates flood damage in an at-risk community through a proactive action that is driven by a flood warning. Therefore, the damage mitigated by the proactive action (D_m^{pro}) only occurs in flooded houses previously warned.

The overestimation of n_{houses}^{warned} (Figure 6.2a) produces D_m^{pro} and a necessary (C_w^n) and/or unnecessary cost of the warning response (C_w^u). C_w^n is the cost associated with warning houses that were subsequently flooded, whereas C_w^u is the cost associated with warned properties that were not subsequently flooded. C_w^u occurs when i) there was a "wrong" decision due to a warning that was issued, and a flood did not occur in the at-risk community (false alarm) (Figure 6.2a2), or ii) there is a 'good' decision, and the magnitude of the forecast flooding was *higher* than the

observed one (Figure 6.2a1). A 'good' decision refers to a situation when a flood in the at-risk community is preceded by a warning (hit). In the case of false alarms, the following condition is fulfilled: $n_{houses}^{flooded} = 0$, $n_{houses}^{warned} > 0$, and $C_w = C_w^u$; whereas in the case of hits: $n_{houses}^{warned} > n_{houses}^{flooded} > 0$, $C_w = C_w^u + C_w^n$, and $D_m^{pro} > 0$ (Figure 6.2a).

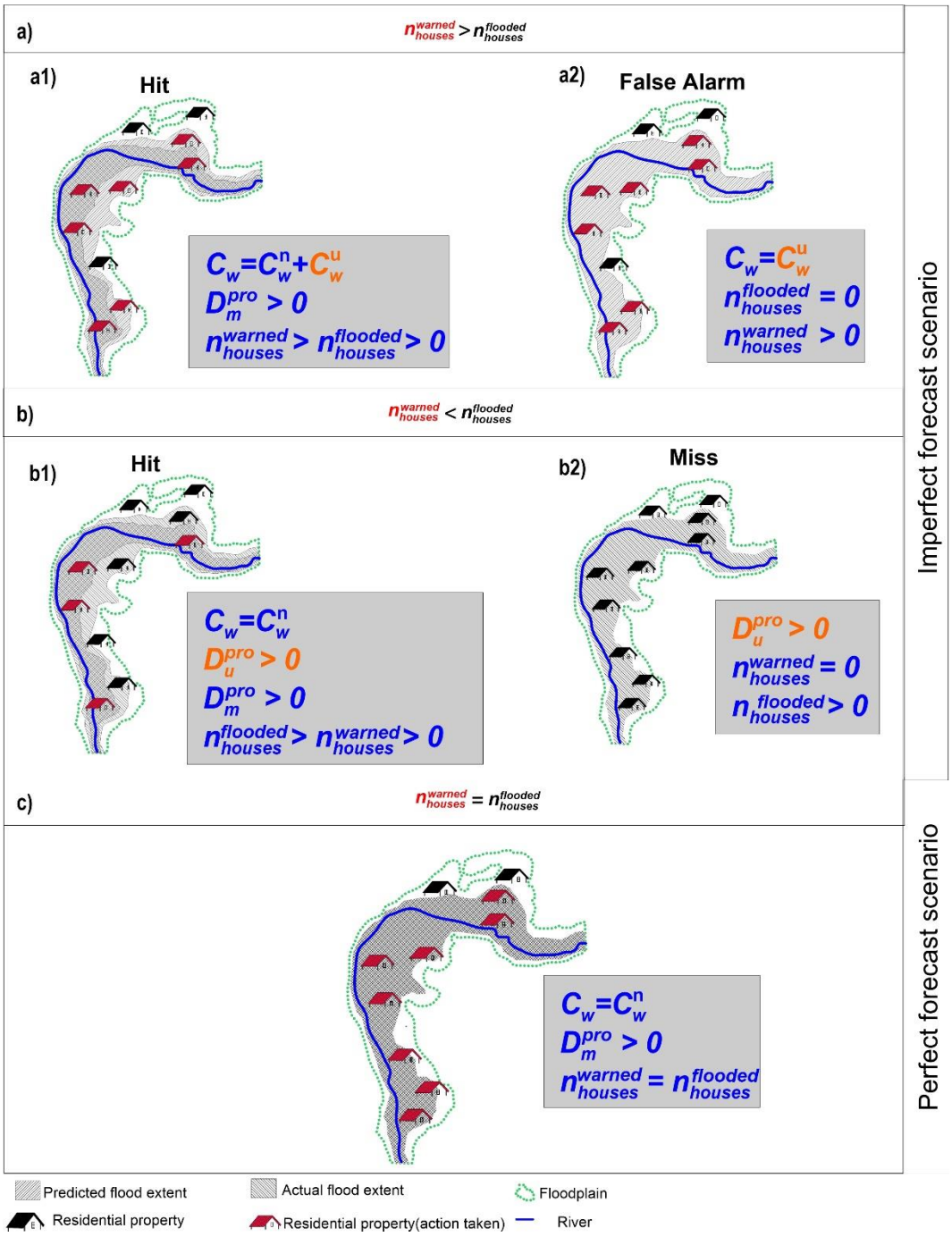


Figure 6.2: Illustration of the net damage associated with forecast uncertainty

Figure a and b illustrate the economic consequences of the four potential situations during the operation of an imperfect FEWS. In these systems, the net damage associate with forecast uncertainty (D_u^{pro} and C_w^u) are present in misses (b2), false alarms (a2), *and* hits (a1 and b1) events. Figure c illustrates the economic consequences in the PFS.

The underestimation of n_{houses}^{warned} (Figure 6.2b) produces D_m^{pro} and C_w^n and/or damage not mitigated by the proactive action (D_u^{pro}). The latter damage occurs when i) there was a "wrong" decision due to a warning *not* being issued, and a flood *occurred* in the at-risk community (miss) (Figure 6.2b2), or ii) there is a hit, and the magnitude of the forecast flooding was *lower* than the observed one(Figure 6.2b1). In the case of misses, the following condition is fulfilled: $n_{houses}^{flooded} > 0$, $n_{houses}^{warned} = 0$, $D_u^{pro} > 0$; in the second case: $n_{houses}^{flooded} > n_{houses}^{warned} > 0$, $D_u^{pro} > 0$, $D_m^{pro} > 0$, and $C_w = C_w^n$ (Figure 6.2b).

Since C_w^u and D_u^{pro} occur when $n_{houses}^{warned} \neq n_{houses}^{flooded}$ which, in turn, only occurs when there is forecast uncertainty, this cost and damage are defined in this work as the *net damage associated with forecast uncertainty*. When analysing the benefits of a FEWS based on imperfect forecasts, this net damage should be considered. The only scenario when this net damage is nil is the PFS. Under this scenario, there is no forecast uncertainty, and n_{houses}^{warned} is neither overestimated nor underestimated, i.e., $n_{houses}^{warned} = n_{houses}^{flooded}$. Therefore, the economic consequences in this scenario are associate with C_w^n and D_m^{pro}(Figure 6.2c). Note that all these variables are a function of the observed or forecasted depths of water causing damage which will be made explicit later, but this dependence is excluded here for simplicity.

To the best of the author's knowledge, a framework that analyses the economic effectiveness of a FEWS in terms of the expected economic benefits by considering the net damage associated with the forecast uncertainty *in all possible situations* (Figure 6.2a and Figure 6.2b) has not been created before. Verkade and Werner (2011) propose a framework to have an approximation of these benefits by combining a hydro-economic model with the theory of relative economic value. However, their framework has two main limiting assumptions:

- ✓ The cost of the warning response is assumed to be independent of the magnitude of the forecast flooding, and it is not estimated as a function of this variate.
- ✓ C_w^u and D_u^{pro} are only estimated as a result of misses and false alarms (Figure 6.2a2 and Figure 6.2b2). Their framework, therefore, negates the fact that the net damage associated with the forecast uncertainty can also be present in a hit event (Figure 6.2a1 and Figure 6.2b1).

The hydro-economic ED model used in this chapter overcomes the assumptions mentioned above. As was mentioned in the introductory part, it is based on a MC simulation framework and can be used to gain insight into i) the estimation and relationship of the economic effectiveness and reliability of a FEWS based on inundation forecasting and ii) the effects of the predictive uncertainty on its economic benefits. The hydro-economic ED model and the main results of this analysis are shown in the following sections.

6.3 Analysis of the flood warning performance

The hydro-economic ED model evaluates the flood warning performance in terms of economic effectiveness and reliability. The metrics used by the framework to represent these two flood warning performance attributes are explained as follows.

6.3.1 Economic effectiveness

The economic effectiveness of a FEWS, E_w, in essence, represents the expected economic benefits associated with a warning service and with a forecasting scenario. It is defined in this work as the economic flood risk reduction relative to the economic flood risk of the NWS (Eq. 6.1). For a totally ineffective FEWS, E_w = 0, but a perfect FEWS will not yield a value of 1 as it can only mitigate a proportion of the damage. The forecasting scenarios can be split into the perfect case, PFS, and imperfect forecast scenarios, i.e., the DFS or PrFS. Since ED_w in Eq. 6.1 includes C_w, this metric represents the expected net damage associated with the warning scenario. Only if E_w>0 will the forecasts of a FEWS have *value or utility* in the at-risk community.

Figure 6.3 illustrates the concept of ED for each scenario considered by the hydro-economic ED model to estimate E_w through Eq. 6.1. In the NWS, a proactive action is not conducted in the at-risk community; ED_{nw}(Figure 6.3a), is, therefore, obtained by integrating a damage function that represents the damage D_u^{pro}. In the PFS, it is assumed that flood damage in the at-risk community is always mitigated by the proactive action, and the net damage associated with forecast uncertainty is nil; the ED of the PFS (ED_w^{pf}) (Figure 6.3b) is ,therefore, obtained by integrating a damage function based on the depth of flooding with the frequency of flooding that represents the net damage D_w^{pf} which includes D_m^{pro} and C_w^n (Figure 6.2c). In

the imperfect forecast scenarios, there is forecast uncertainty and, therefore, $n_{houses}^{warned} \neq n_{houses}^{flooded}$; the ED of an imperfect scenario (ED_w^{if}) (Figure 6.3c) is , therefore, obtained by integrating a bivariate damage function that represents the net damage D_w^{if} which includes all the economic consequences illustrated in Figure 6.2a and Figure 6.2b. In an imperfect forecast scenario, the net damage associated with forecast uncertainty (C_w^u and D_u^{pro}) impact negatively on ED_w^{if} by increasing its value. Since this net damage is nil in the PFS (Figure 6.2c), the value ED_w^{pf} is always lower than the values obtained based on an imperfect forecast scenario. That also means that the value of E_w (Eq. 6.1) based on the PFS (E_w^{pf}) represents the maximum economic benefits that can be achieved by the system; the economic benefits of an imperfect forecast scenario, i.e., the DFS or PrFS, cannot overcome E_w^{pf}.

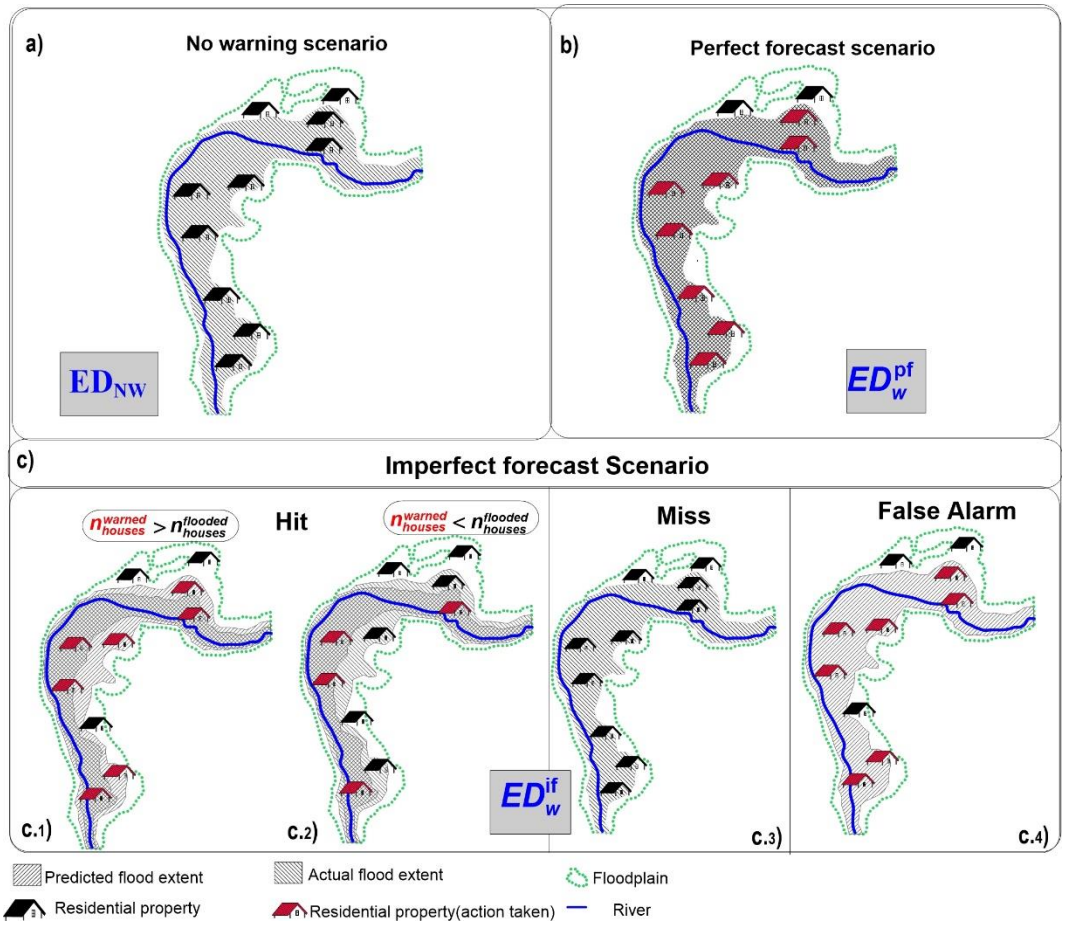


Figure 6.3: Illustration of the estimation of ED for each scenario considered by the hydro-economic ED model.

The ED is obtained by integrating a damage function that represents the damage D_u^{pro}, D_w^{pf} and D_w^{if} in the NWS, PFS, and imperfect forecast scenario, respectively. The net damage D_w^{pf} includes D_m^{pro} and C_w^n (Figure 6.2c), and D_w^{if} includes all the economic consequences illustrated in Figure 6.2a and Figure 6.2b.

6.3.2 Reliability

To estimate the ED in the imperfect forecast scenarios, one should represent, for each potential flood in the at-risk community, the flood warning decision which, for the case of an inundation map forecast, will be the number of houses warned, n_{houses}^{warned} , and the outcome will be the number of flooded houses, $n_{houses}^{flooded}$. With that information, one can also evaluate the reliability of a FEWS in terms of whether the warned properties were or were not subsequently flooded. This criterion was called in Chapter 5 floodplain property-based criterion (FPC). Section 5.4.2 explains how the FPC defines hits, misses, and false alarms events in terms of affected houses to evaluate the flood warning reliability, whereas section 5.5 explains how the metrics known as POD and FAR can be used to aggregate the outcomes of the FPC into a single value to have a snap-shot of the reliability of the FEWS. In essence, the FPC uses the POD and FAR to answer the following questions: What is the probability of a flooded house being correctly warned? What is the probability of a warned house being incorrectly warned? (Table 2.2). The hydro-economic ED model uses these concepts to evaluate the reliability of the inundation-forecasting based FEWS in the DFS and PrFS cases.

6.4 Main components of the hydro-economic ED model

The hydro-economic ED model has three main components: the MCFG, the FWDC, and the RIC (Figure 6.1a). Each of them is introduced in the following subsections.

6.4.1 The Monte Carlo flood and forecast generator (MCFG)

The hydro-economic ED model uses a MCFG to generate bivariate values of observed peak water levels y and their forecasts $\hat{y}$, where y defines the magnitude of potential floods in the at-risk community. The MCFG uses a bivariate probability distribution for this purpose. The adopted bivariate model in the MCFG was the bivariate Lognormal distribution, and the bivariate simulation is done based on the algorithm shown in section 4.4.4. The marginal distribution $f(y)$ of the bivariate distribution $f(y, \hat{y})$ is used to represent the magnitude of the potential floods in the at-risk community in the NWS and PFS. This component also uses a lead time-performance function (Figure 4.14) to assign to a lead time τ a value of $\rho_{y\hat{y}}$.

6.4.2 The flood warning decision-making component (FWDC)

The flood warning decision-making component FWDC is used in the hydro-economic ED model to simulate warning decisions in the DFS and PrFS. These decisions are represented by decision rules and consider deterministic and probabilistic forecasts, respectively. In the DFS, the deterministic forecast-based flood-warning decision is represented by the decision rule DFDR explained in section 5.6.1, which assumes that a warning is issued when $\hat{y}$ is greater than the flooding threshold (y_T) of the at-risk community defined by the Standard of Protection (SoP)(Figure 5.5b). In the PrFS, the probabilistic forecast-based flood warning decision is represented by the decision rule PDR explained in section 5.6.2. This rule uses a probabilistic threshold PT to define a warning level $\hat{y}_w$ from $f(y|\hat{y})$ at which the Warner should base his/her decisions, where PT, in essense, represents the probability of exceedance of $\hat{y}_w$. Then, the warning decision is based on the direct comparison between $\hat{y}_w$ and the flooding threshold y_T (Figure 5.6b). In the hydro-economic model, it is assumed that, in the PDR case, the Warner acts to *increase* E_w (decrease the flood risk) in the at-risk community. The aim is, thus, to find an optimal probabilistic threshold (PT^*) that fulfils this purpose. Note that PT^* in this case will be different from that obtained in Chapter 5 which used the maximum value of POD minus FAR as the optimization criterion

6.4.3 The response and impact component (RIC)

The response and impact component RIC is used by the hydro-economic ED model to represent the flood damage in the floodplain without the warning service and net flood damage with the warning service. Here, the impact is only measured in terms of damage to house contents, and no attempt is made to represent psychological damage to residents, or in the worst case, loss of life. These damages are represented by this component through several expressions that depend on the variates describing damage considered in this work (flood depth and extension). In this section, the development of these equations and the rationality behind them are presented. These equations are used to build the damage functions used to estimate the ED of the scenarios shown in Figure 6.3. Before describing them, the section starts by explaining how the damage variates needed to use these equations are simulated in this component of the hydro-economic ED model.

6.4.3.1 Simulation of observed and forecast floods

To represent the flood damage in the floodplain without and with the provision of a FEWS, one must simulate the observed and forecast inundation in the at-risk community. The damaging variates considered in this work are the flood depth and extension. The RIC does not simulate these variates directly but simulates their target information. The target information in such an analysis is i) n_{houses}^{warned} and $n_{houses}^{flooded}$ and ii) the flood depth reached in each of them. The former is represented in this work through the impact curve-based method used in Chapter 5 (Figure 5.3). To estimate the flood depth, the RIC assumes that the elevation of the floodplain properties with respect to the ground is 0.25 m (Figure 6.4). It is also assumed that the water surface elevation at a floodplain property is not significantly different from the water surface elevation in the channel. In this case, the flood depth can be estimated as the difference between the magnitude of the flooding and the bankfull level ($y_{bankfull}$), once the water level has overtopped the flooding threshold (y_T) defined by the SoP, and water spills into the floodplain. A small allowance is made for the level of the property floors above $y_{bankfull}$. Section 4.5.3 (Eq. 4.54) explains how y_T is computed based on the assumed SoP. The bankfull condition is assumed in this work to be defined by a 2.5-year return period

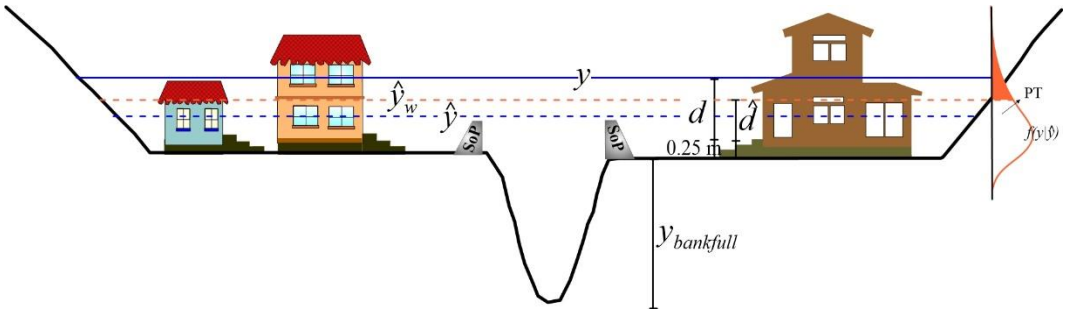


Figure 6.4: Illustration of the flood depth estimation in the RIC for an observed flood and its forecast

The RIC assumes that the elevation of the floodplain properties with respect to the ground is 0.25 m. The figure illustrates the case when the observed peak level y and its forecast, represented by $\hat{y}$ and $\hat{y}_w$ in the DFS and PrFS, respectively, overtop the flooding threshold y_T defined by the standard of protection (SoP).

Under these assumptions, the RIC simulates the target information of the observed and forecast flood inundation through two basic steps:

- ✓ ***Step 1:*** Estimate n_{houses}^{warned} and $n_{houses}^{flooded}$ through the impact curve-based method explained in section 5.3 (Figure 5.3). n_{houses}^{warned} is estimated based on the variate on which the warning decision is based (ν). In the DFS, ν

represents $\hat{y}$ (Figure 5.5b), whereas, in the PrFS, ν represents a warning level $\hat{y}_w$ estimated from $f(y|\hat{y})$ (Figure 5.6b).

- ✓ ***Step 2***: Estimate the observed flood depth and its forecast as:

$$d = y - (y_{bankfull} + 0.25) \qquad \text{Eq. 6.2}$$

$$\hat{d} = \nu - (y_{bankfull} + 0.25) \qquad \text{Eq. 6.3}$$

where d and $\hat{d}$ is the observed flood depth and its forecast estimated as a function of y and ν, respectively.

By conducting the two steps mentioned above, the RIC simulates, for each observed or forecast flooding event (defined as the values of y or ν greater than y_T), n_{houses}^{warned}, $n_{houses}^{flooded}$, $\hat{d}$ and d, which is the information needed to use the expressions that represent the flood damage and net flood damage in the floodplain without and with the flood warning service, respectively. They are described in the next subsections.

6.4.3.2 The damage without the warning service

The flood damage in an at-risk community depends on the flooding characteristics and the vulnerability (the lack of resistance to damaging forces), and exposure of the floodplain assets. Without the warning service, this exposure and vulnerability depend on the standard of protection, SoP, and the vulnerability and exposure of the floodplain assets. In this scenario, the SoP (structural measures such as river banks, flood dykes, and dams) reduces but does not eliminate the flood risk (Figure 6.4), and no proactive action is taken to reduce that residual flood risk because there is no warning service. To represent the flood damage, the RIC considers only the direct economic damage, and the floodplain assets considered are only the contents of properties. An expression that represents the direct economic damage to these contents for every flood in the at-risk community can be expressed by:

$$D_u^{pro}(d) = \sum_{k=1}^{n_{houses}^{flooded}} \vartheta_k \, D_k^c(d_k) \qquad \text{Eq. 6.4}$$

where D_u^{pro} and $n_{houses}^{flooded}$, as explained above, represent the damage in the floodplain unmitigated by a proactive action (without the warning service) and the

number of flooded houses, respectively, ϑ_k is the monetary value of the content of the flooded property k, D_k^c, expressed as a percentage, is the damage to this content without warning service as a function of the flood depth, and d_k is the flood depth at the flooded property k.

The number of flooded houses is expressed in Figure 5.7 as a function of y, and therefore d in Eq. 6.2. This function assumes a gently varying terrain where the number of flooded houses increases with y and therefore d. If the terrain in the floodplain is complex because of variable topography, the estimation of D_u^{pro} may need hydraulic inundation modelling to represent d_k at each flooded property. However, this work is limited to the analysis of floodplains with gently varying terrain, where d_k can be assumed to be the same in each flooded property and can be considered as an average depth across the number of houses flooded (Figure 6.4). In addition, if one assumes an average monetary value of the residential content in each floodplain property, D_u^{pro} can be expressed by:

$$D_u^{pro}(d) = \vartheta D_c(d) n_{houses}^{flooded} \qquad \text{Eq. 6.5}$$

where ϑ is the average monetary value of the residential content in the floodplain properties, and $D_c(d)$, expressed as a percentage, is the damage to this content without a proactive action as a function of the flood depth d in the floodplain computed based on Eq. 6.2. The average monetary value ϑ is an input parameter of the framework and the value of $D_c(d)$ is obtained from the family of damage curves proposed by Carsell *et al.* (2004), which represent the percentage damage to the residential content as a function of the mitigation time and flood depth (Figure 6.5). In this thesis, it is assumed that the mitigation time is equal to the lead time τ, and $D_c(d)$ is represented by damage to the content when the (mitigation) lead time is zero, i.e., $D_c^{\tau=0}$. The assumptions made to estimate $D_u^{pro}(d)$ through Eq. 6.5 reduce considerably the computational resources required to estimate the flood damage.

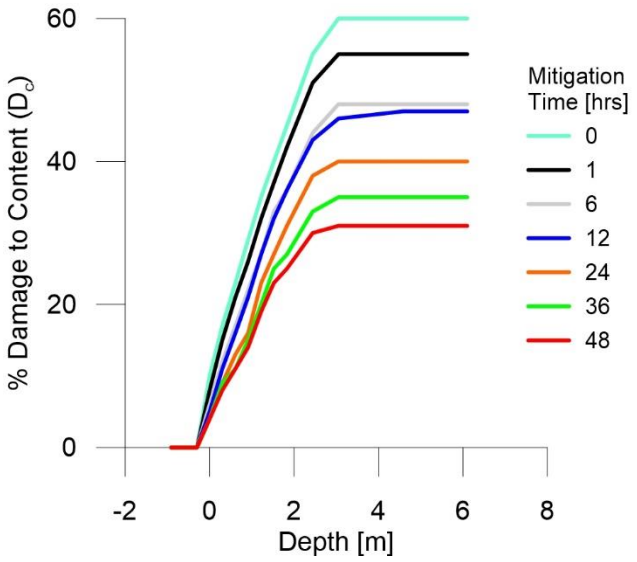


Figure 6.5: The residential content depth-percentage damage relationship with flood mitigation time

Source: Carsell *et al.*, (2004).

Eq. 6.5 is used by the hydro-economic ED model to represent the flood damage in the NWS (Figure 6.3a) and to represent the flood damage of misses (Figure 6.2b2) in the imperfect forecast scenarios.

6.4.3.3 The damage with the proactive action

A FEWS mitigates the residual risk in the at-risk community associated with an SoP through the proactive action conducted in the warning response. The proactive action can modify i) the exposure probability of persons or mobile items or ii) if the lead time is long enough, the vulnerability of the structural assets through the implementation of supplementary intervention measures, e.g., demountable defences. The proactive action considered in the RIC has to do with *moving and/or evacuating the contents* of the floodplain properties after a flood warning is issued. Taking into account the same assumptions made to estimate $D_u^{pro}(d)$ through Eq. 6.5, an expression that represents the potential damage avoided in the floodplain when that proactive action is conducted can be defined by:

$$D_p^{\tau}(d) = \vartheta[D_c(d) - D_c^{\tau}(d)]n_{houses}^{flooded} \qquad \text{Eq. 6.6}$$

where $D_p^{\tau}(d)$ represents the potential damage avoided by the proactive action conducted after a flood warning within a lead time τ was issued and when there is a flood depth d in the at-risk community, and $D_c^{\tau}(d)$ is the percentage damage to

residential contents after the proactive action has been conducted, which is obtained from the functions shown in Figure 6.5.

$D_p^\tau(d)$ can be affected when householders i) do not heed the warning and act effectively or ii) are involved in evacuation decisions when the flood is not shallow, and their lives are at risk, and iii) are not notified in time. To account for this, $D_p^\tau(d)$ is multiplied by an economic efficiency parameter α which ranges from 0 to 1 and represents the uncertainty of the proactive action. It is analogue to those proposed by Parker (1991) and Carsell *et al.* (2004), as it decreases the potential damage avoided to obtain the value known as the actual flood damage avoided. This avoidable damage can be expressed by:

$$D_a^\tau(d) = \alpha D_p^\tau(d) \qquad \text{Eq. 6.7}$$

where α is an input parameter of the hydro-economic ED model and $D_a^\tau(d)$ represents the actual flood damage avoided by the proactive action conducted after a flood warning with a lead time τ was issued and a flood depth d occurred in the at-risk community. According to Eq. 6.7, $D_a^\tau(d)$ increases when d and $n_{houses}^{flooded}$ or τ increases. Thus it represents well the damage saved by FEWSs, which is likely to be greater for high than for low flood stages and increase as the lead time increases (Carsell *et al.*, 2004).

The flood damage in the floodplain mitigated by the proactive action can be computed by subtracting $D_a^\tau(d)$ from $D_u^{pro}(d)$(Eq. 6.5).

$$D_m^{pro,\tau}(d) = D_u^{pro}(d) - D_a^\tau(d) \qquad \text{Eq. 6.8}$$

where $D_m^{pro,\tau}(d)$ represents the flood damage in the floodplain mitigated by the proactive action (moving and or evacuating the residential contents) conducted after a flood warning with lead time τ was issued. If $D_u^{pro}(d)$ and $D_a^\tau(d)$ are replaced in Eq. 6.8 by Eq. 6.5 and Eq. 6.7 and Eq. 6.6, respectively, $D_m^{pro,\tau}(d)$ can be expressed by:

$$D_m^{pro,\tau}(d) = \vartheta n_{houses}^{flooded}\{D_c(d) - \alpha[D_c(d) - D_c^\tau(d)]\} \qquad \text{Eq. 6.9}$$

Eq. 6.9 is used by the hydro-economic ED model to represent the mitigated flood damage in the at-risk community in the PFS case (Figure 6.3b) and to represent the mitigated flood damage when n_{houses}^{warned} is overestimated in a hit event (Figure 6.2a1). The latter is possible because, in this situation, all the flooded houses are warned.

6.4.3.4 The damage with and without the proactive action

The expressions used by the hydro-economic ED model to represent the flood damage of most of the potential situations in an imperfect forecast scenario have already been explained in the prior subsections. There is only one type of flood damage associated with these potential situations whose equation has not been described yet. It has to do with the representation of the flood damage when n_{houses}^{warned} is underestimated in a hit event (Figure 6.2b1). In this case, the flood damage has to be split into flood damage mitigated and not mitigated by the proactive action, respectively, and can be described by:

$$D_{u,m}^{pro,\tau}(d) = \vartheta D_c(d) n_{houses}^{flooded,nw} + \vartheta D_c^{\tau}(d) n_{houses}^{flooded,w} \qquad \text{Eq. 6.10}$$

where $D_{u,m}^{pro,\tau}(d)$ represents the flood damage mitigated and unmitigated by the proactive action in the at-risk community in a hit event when $y < v$, i.e. $n_{houses}^{warned} < n_{houses}^{flooded}$, and $n_{houses}^{flooded,w}$ and $n_{houses}^{flooded,nw}$ represent the flooded houses warned and no warned, respectively. The other parameters have already been introduced above. Note that in this situation $n_{houses}^{flooded,w} = n_{houses}^{warned}$ and $n_{houses}^{flooded,nw} = n_{houses}^{flooded} - n_{houses}^{warned}$ (Figure 6.2b1).

6.4.3.5 The cost of the warning response

A flood warning system mitigates the residual risk associated with an SoP through the warning response, which can only be made at a cost C_w. As was explained in section 6.2, this cost refers to the cost of issuing a warning and any cost incurred in the dissemination process and by the public taking action (proactive action). That cost is driven by the forecast flood because it is the only information available when the warning is issued. The cost of the warning response C_w is here assumed to be a continuous variate proportional to the economic benefits (damage saved) of the FEWS during a flood event. Since the avoidable damage in the at-risk community

(Eq. 6.7), in essence, represents those economic benefits, the RIC uses $D_a^\tau(d)$ to estimate C_w, in the PFS case as:

$$C_w^\tau(d) = \gamma D_a^\tau(d) \qquad \text{Eq. 6.11}$$

where $C_w^\tau(d)$ is the cost of the warning response associated with a lead time τ, γ is a parameter ($0 < \gamma < 1$) used to represent $C_w^\tau(d)$ as a proportion of $D_a^\tau(d)$. If $D_a^\tau(d)$ is replaced in Eq. 6.11 by Eq. 6.7 and Eq. 6.6, $C_w^\tau(d)$ can be expressed by:

$$C_w^\tau(d) = \gamma\alpha\vartheta[D_c(d) - D_c^\tau(d)]n_{houses}^{flooded} \qquad \text{Eq. 6.12}$$

The parameter γ is an input parameter of the hydro-economic ED model, and it uses Eq. 6.12 to represent $C_w^\tau(d)$ in the PFS case. Note that $C_w^\tau(d)$ in that equation is a function of $n_{houses}^{flooded}$ because in the PFS it is assumed $n_{houses}^{warned} = n_{houses}^{flooded}$, and, therefore, $C_w^\tau(d) = C_w^n(d)$ (see Figure 6.2c to see the definition of $C_w^n(d)$). That means, in essence, that we are assuming perfect knowledge about the flooding characteristics. However, in the imperfect forecast scenarios, the only information available when the warning is issued is its forecast. In this case, $n_{houses}^{warned} \neq n_{houses}^{flooded}$, and the RIC first estimates the economic benefits of the potential flood as a function of ν, i.e., $\hat{y}$ or $\hat{y}_w$ for the DFS or PrFS, respectively, and then, it uses γ to compute $C_w^\tau(\hat{d})$ as a percentage of that benefits. This process is represented by simply replacing in Eq. 6.12 $n_{houses}^{flooded}$ and d by n_{houses}^{warned} and $\hat{d}$, respectively, to give:

$$C_w^\tau(\hat{d}) = \gamma\alpha\vartheta\left[D_c(\hat{d}) - D_c^\tau(\hat{d})\right]n_{houses}^{warned} \qquad \text{Eq. 6.13}$$

where $\hat{d}$, as was explained above, is computed through Eq. 6.3. Since $C_w^\tau(\hat{d})$ in Eq. 6.13 is proportional to n_{houses}^{warned}, the RIC model uses that expression to represent $C_w^n(\hat{d})$ and $C_w^u(\hat{d})$ in all possible conditions in an imperfect forecast scenario (see Figure 6.2). Thus, in a false alarm event (Figure 6.2a2), $C_w^\tau(\hat{d})$ respresents $C_w^u(\hat{d})$. When n_{houses}^{warned} is overestimated in a hit event (Figure 6.2a1), $C_w^\tau(\hat{d})$ represents $C_w^n(\hat{d}) + C_w^u(\hat{d})$. Finally, when n_{houses}^{warned} is underestimated in a hit event (Figure 6.2b1), $C_w^\tau(\hat{d})$ respresents $C_w^n(\hat{d})$.

Note that according to Eq. 6.12 and Eq. 6.13, $C_w^\tau(d)$ and $C_w^\tau(\hat{d})$ in the perfect and imperfect forecast scenario, respectively, increases as the affected houses and flood depth and τ increase. Therefore, these expressions represent the behaviour of the cost of the warning response in FEWSs as the higher the level of damage, the higher the spending needed to mitigate this damage. Furthermore, the cost of the warning response may increase as the lead time increases (Verkade and Werner, 2011; Matte *et al.*, 2017).

6.4.3.6 The net damage with the perfect warning service (perfect forecast scenario)

Flood damage in the floodplain with the warning service should include the mitigated damage by the proactive action *and* the cost associated with the warning response associated with a lead time τ, i.e., $D_m^{pro,\tau}(d)$ and $C_w^\tau(d)$, respectively. The net damage in the PFS can be, therefore, represented by:

$$D_w^{pf,\tau}(d) = D_m^{pro,\tau}(d) + C_w^\tau(d) \qquad \text{Eq. 6.14}$$

where $D_w^{pf,\tau}(d)$ represents the economic net damage in the floodplain associated with a flood depth d, and the warning service with a given lead time τ, and considering a PFS. If one replaces in Eq. 6.14 $D_m^{pro,\tau}(d)$ and $C_w^\tau(d)$ by Eq. 6.9 and Eq. 6.12, $D_w^{pf,\tau}(d)$ can be represented by:

$$D_w^{pf,\tau}(d) = \vartheta n_{houses}^{flooded}\{D_c(d) - \alpha(1-\gamma)[D_c(d) - D_c^\tau(d)]\} \qquad \text{Eq. 6.15}$$

Eq. 6.15 is used by the hydro-economic ED model to represent the net damage in the PFS (Figure 6.2c).

6.4.3.7 The net damage with the imperfect warning service

The net damage of an imperfect warning service $D_w^{if,\tau}(d)$ is simulated through several expressions described in the prior subsections, which represent the economic consequences of all potential situations when $n_{houses}^{warned} \neq n_{houses}^{flooded}$ (Figure 6.2a and Figure 6.2b). Thus, the different values that $D_w^{if,\tau}(d)$ can take are dependent on whether the outcome of the warning is a hit, miss, or false alarm

(*defined in terms of flood warning decisions*). The expressions for each case are tabulated in Table 6.1 and are explained as follows.

As was explained in section 6.2, a hit in terms of a warning decision is referred to as the situation when a flood in the at-risk community is preceded by a warning. These events produce mitigated flood damage because the proactive action was conducted between the issue of the warning and the onset of the flood with magnitude y. These events can also produce unmitigated damage when n_{houses}^{warned} is underestimated (Figure 6.2b1). The damage reduction is achieved at a cost $C_w^{\tau}(\hat{d})$ estimated in this work as a function of v (and $\hat{d}$ (Eq. 6.3)), which, as was explained above, represents $\hat{y}$ and $\hat{y}_w$ in the DFS and PrFS, respectively. Thus, the net damage for any pair (y, v) defined as a hit where $y < v$ (these events are noted in this work as hits 1), i.e., $n_{houses}^{flooded} < n_{houses}^{warned}$ (Figure 6.2a1) is represented by $D_m^{pro,\tau}(d)$ plus $C_w^{\tau}(\hat{d})$, Eq. 6.9 and Eq. 6.13, respectively. If $y > v$ (these events are noted in this work as hits 2), i.e., $n_{houses}^{flooded} > n_{houses}^{warned}$ (Figure 6.2b1), this net damage is represented by $D_{u,m}^{pro,\tau}(d)$ plus $C_w^{\tau}(\hat{d})$, Eq. 6.10 and Eq. 6.13, respectively.

Table 6.1: Description of the equations used by the hydro-economic model to represent the economic consequences of an imperfect flood warning system

Event	Economic consequence	Equation
Hits 1 ($y < v$)	$D_m^{pro,\tau}(d) + C_w^{\tau}(\hat{d})$	$\vartheta n_{houses}^{flooded}\{D_c(d) - \alpha[D_c(d) - D_c^{\tau}(d)]\} + r\alpha\vartheta\left[D_c(\hat{d}) - D_c^{\tau}(\hat{d})\right]n_{houses}^{warned}$
Hits 2 ($y > v$)	$D_{u,m}^{pro}(d) + C_w^{\tau}(\hat{d})$	$\left[\vartheta D_c(d) n_{houses}^{flooded,nw} + \vartheta D_c^{\tau}(d) n_{houses}^{flooded,w}\right] + r\alpha\vartheta\left[D_c(\hat{d}) - D_c^{\tau}(\hat{d})\right]n_{houses}^{warned}$
False alarms	$C_w^{\tau,fa}(\hat{d})$	$r\alpha\vartheta\left[D_c(\hat{d}) - D_c^{\tau}(\hat{d})\right]n_{houses}^{warned}$
Misses	$D_u^{pro,\tau}(d)$	$\vartheta D_c(d) n_{houses}^{flooded}$
Correct negative	0	-

Missed events in terms of warning decisions are defined when a flood event occurred in the at-risk community and a warning was not issued (Figure 6.2b2). These events result in unmitigated damage because a proactive action was not conducted, and, therefore, $C_w^{\tau}(\hat{d})$=0. For any pair (y, v) defined as a miss, the damage of misses is, therefore, represented by $D_u^{pro,\tau}(d)$(Eq. 6.5). If the flood event did not occur and a warning was issued, this is considered a false alarm (Figure 6.2a2). In this case, there is no damage in the at-risk community, and the warning

response was conducted in vain; thus, the economic consequences for any pair (y, v) defined as false alarm is represented by $C_w^{\tau,fa}(\hat{d})$ (Eq. 6.13). Finally, if the flood event did not occur, and a warning was not issued, this is considered a correct negative. If the operational costs are excluded from the analysis, this situation does not represent any cost to the system.

6.5 Flood risk model

In general terms, flood risk is defined as the combination of the probability of an event and its negative consequences associated with damage to human health and life, the environment, and economic activity (European Parliament, 2007). The flood risk of a flood-prone area is defined by a metric that in essence aggregates all the potential damages into a single value. It can be estimated quantitatively if the losses are measurable, e.g., monetary or loss of life units, or qualitatively, e.g., allocation in classes, in the case of intangible damages associated with the social, environmental, and cultural impacts (Laoupi and Tsakiris, 2007). Since this thesis concentrates on economic damage only, in this section, the concept of flood risk is used by the model to aggregate all the potential economic damages or net damage values into a single value for all scenarios considered. That single value is the ED, which is the target information when estimating E_w (Eq. 6.1). Thus, this section explains the computation of the ED for each scenario.

6.5.1 The no warning scenario

Flood risk is usually defined as the product of the probability of flooding and the associated consequences, integrated over all possible events. The expected damage ED is the metric traditionally used to define the economic flood risk and give us a snap-shot of flood risk at present or in a future scenario in the at-risk community.

The ED for the NWS can be defined by:

$$ED_{nw} = \int D_u^{pro}(d) f(d) dd \qquad \text{Eq. 6.16}$$

where ED_{nw}, as was explained in section 6.3.1, is the ED of the at-risk community in the NWS (Figure 6.3a), $D_u^{pro}(d)$ is a damage function expressed as a function of d (Eq. 6.5), and $f(d)$ is the probability density function of this variate.

6.5.2 The perfect forecast scenario

If the FEWS is based on perfect forecasts, a flood event is always preceded by a warning, and there is no uncertainty in the inundation forecasts. Thus, the reliability of the flood warning system is assumed perfect, i.e., every flood is preceded by a warning, and every warned property is then flooded (Figure 6.2c). The flood risk in the PFS is defined by:

$$ED_w^{pf,\tau} = \int D_w^{pf,\tau}(d) f(d) dd \qquad \text{Eq. 6.17}$$

where $ED_w^{pf,\tau}$ is the ED of the at-risk community in the PFS (Figure 6.3b), and $D_w^{pf,\tau}(d)$ is a damage function expressed as a function of d (Eq. 614). Since the damage saved by the proactive action depends on the lead time τ, there is a specific value of $ED_w^{pf,\tau}$ for each τ.

6.5.3 The imperfect forecast scenario

The ED of the at-risk community with the warning service based on imperfect forecasts is defined by:

$$ED_w^{if,\tau} = \iint D_w^{if,\tau}(d, \hat{d}) f(d, \hat{d}) dd d\hat{d} \qquad \text{Eq. 6.18}$$

where $ED_w^{if,\tau}$ is the ED of the at-risk community with the warning service assuming an imperfect forecast scenario (Figure 6.3c), and $D_w^{if,\tau}(d, \hat{d})$ is a bivariate net damage function that describes the economic consequences of hits, misses, and false alarms (*defined in terms of flood warning decisions*) according to Table 6.1, and $f(d, \hat{d})$ is the joint distribution of d and $\hat{d}$. Since $D_w^{if,\tau}(d, \hat{d})$ represents the net damage of a FEWS with a given lead time τ, and the economic consequences depend on the warning decision, there is a specific value of ED_w^{if} for a given warning decision rule and lead time τ.

6.5.4 Estimation of ED

The analytical expression of the ED (economic flood risk) for each warning scenario considered in the analysis was defined in the prior subsections. To estimate it, one must solve the integrals shown in Eq. 6.16, Eq. 6.17, and Eq. 6.18. An

estimation of these integrals can be obtained by first building the functions that describe the damage or net damage as described in section 6.4 above and then solving the integrals via an analytical solution or numerical integration (Olsen *et al.*, 2015). In this framework, the latter approach was used. Based on this approach, the integral of the NWS (Eq. 6.16) can be estimated by:

$$\widehat{ED}_{nw} = \frac{1}{n}\sum_{i=1}^{K} D_{u,i}^{pro}(d_i) f_i(d_i) \qquad \text{Eq. 6.19}$$

where n represents the number of values sampled from the marginal distribution of d, K is the number of intervals considered in the numerical integration defined by an increment of integration Δd, d_i is the flood depth located at the midpoint of the interval i, $D_{u,i}^{pro}(d_i)$ is the unmitigated damage associated with d_i, and $f_i(d_i)$ is the number of sampled values falling in the interval i.

A similar approach is used to estimate the integral of the perfect forecast scenario (Eq. 6.17) as:

$$\widehat{ED}_{w}^{pf,\tau}(\tau) = \frac{1}{n}\sum_{i=1}^{K} D_{w,i}^{pf,\tau}(d_i) f_i(d_i) \qquad \text{Eq. 6.20}$$

where $D_{w,i}^{pf,\tau}(d_i)$ is the PFS-based net damage associated with d_i.

The estimate of the double integral of the imperfect forecast scenario (Eq. 6.18) can be obtained by dividing the bivariate space into four quadrants corresponding to hits, misses, false alarms, and correct negatives (*defined in terms of flood warning decisions*), and applying the above estimation approach to each expression of Table 6.1. Therefore, the total expected damage will be:

$$\widehat{ED}_w^{if,\tau} = \frac{1}{n}\left[\sum_{i=1}^{K} D_{m,i}^{pro,\tau}\left(d_i^{h1}\right) f_i\left(d_i^{h1}\right) + \sum_{i=1}^{K} C_{w,i}^{\tau,h1}\left(\hat{d}_i^{h1}\right) f_i\left(\hat{d}_i^{h1}\right) + \sum_{i=1}^{K} D_{um,i}^{pro,\tau}\left(d_i^{h2}\right) f_i\left(d_i^{h2}\right) + \sum_{i=1}^{K} C_{w,i}^{\tau,h2}\left(\hat{d}_i^{h2}\right) f_i\left(\hat{d}_i^{h2}\right) + \sum_{i=1}^{K} C_{w,i}^{\tau,fa}\left(\hat{d}_i^{fa}\right) f_i\left(\hat{d}_i^{fa}\right) + \sum_{i=1}^{K} D_{u,i}^{pro,\tau}\left(d_i^{m}\right) f_i\left(d_i^{m}\right)\right] \quad \text{Eq. 6.21}$$

where the superscript h_1 and h_2 are used to reference the economic variates and observed and forecast flood depths associated with hits 1 and hits 2, respectively, whereas for false alarms and misses, the superscripts fa and m are used for these purposes. The damage variables in Eq. 6.21 are taken from Table 6.1 for the four cases.

6.6 Algorithms used to define the reliability and effectiveness of the flood warning system

The hydro-economic ED model is used in this work to explore the reliability and economic effectiveness of a FEWS based on inundation forecasting. In the prior sections, the components, metrics, and concepts used by the model to define these two flood warning attributes were explained. This section of the chapter describes the algorithms used by the hydro-economic ED model to estimate the ED for the three warning scenarios considered in section 6.5.

The bivariate Lognormal distribution (BLND) is used for all the sampling/sensitivity experiments carried out in this chapter. The economic effectiveness of the FEWS for each with-warning scenario is computed relative to the ED of the NWS, which is described by Eq. 6.19 and computed as follows.

- ✓ ***Step 1***: Use the MCFG to sample n values from $f(y)$.
- ✓ ***Step 2:*** For each sampled value y_i, use the RIC to estimate i) the number of houses flooded $n_{houses}^{flooded,i}$, through the impact-curve base method, and the flood depth d_i by applying Eq. 6.2 (section 6.4.3.1); and ii) $D_u^{pro}(d_i)$, i.e., the flood damage in the at-risk community without the warning service, by applying Eq. 6.5
- ✓ ***Step 3:*** Estimate ED_{nw} (Figure 6.3a) by applying Eq. 6.19.

6.6.1 Perfect forecast scenario

The ED of the PFS, which represents the maximum economic benefits that can be achieved by a FEWS, is estimated by following the algorithm.

- ✓ ***Step 1***: Use the MCFG to sample n values from $f(y)$.
- ✓ ***Step 2***: For each sampled value y_i, use the RIC to estimate i) the number of houses flooded $n_{houses}^{flooded,i}$, through the impact-curve base method, and the flood depth d_i by applying Eq. 6.2 (section 6.4.3.1).
- ✓ ***Step 3:*** For each d_i, uses the RIC to compute $D_w^{pf,\tau}(d_i)$, i.e., the net damage in the at-risk community with the warning service based on perfect forecast, by applying Eq. 6.15.
- ✓ ***Step 4:*** Estimate $ED_w^{pf,\tau}$(Figure 6.3b) through Eq. 6.20.
- ✓ ***Step 5:*** Compute the economic effectiveness of the FEWS by using Eq. 6.1

6.6.2 Imperfect deterministic forecast scenario

The DFS assumes that the warning decision and the inundation forecasting are based only on $\hat{y}$. The algorithm used to estimate the metrics that define the reliability and economic effectiveness under this scenario is detailed as follows.

- ✓ ***Step 1***: Uses the MCFG to sample n bivariate values from $f(y, \hat{y})$ based on the BLND algorithm described in section 4.4.4.
- ✓ ***Step 2:*** For each forecast value $\hat{y}_i$, uses the FWDC to simulate the warning decision according to the DFDR (Eq. 5.1), i.e., warn only if $\hat{y}_i > y_T$.
- ✓ ***Step 3:*** Use the flooding threshold criterion (FTC) described in Chapter 5 to define each pair $(y_i, \hat{y}_i)$ as $(y_i^h, \hat{y}_i^h)$, $(y_i^m, \hat{y}_i^m)$, and $(y_i^f, \hat{y}_i^f)$, i.e., as hit, miss, and false alarm events, respectively (Table 5.2). In addition, define each pair $(y_i^h, \hat{y}_i^h)$ as $(y_i^{h1}, \hat{y}_i^{h1})$ and $(y_i^{h2}, \hat{y}_i^{h2})$, i.e., hits 1 and hits 2, respectively.
- ✓ ***Step 4:*** Use the RIC to define the economic consequences of each pair $(y_i^{h1}, \hat{y}_i^{h1})$, $(y_i^{h2}, \hat{y}_i^{h2})$,$(y_i^m, \hat{y}_i^m)$, and $(y_i^f, \hat{y}_i^f)$ according to Table 6.1.
- ✓ ***Step 5:*** Estimate the ED for the DFS ($\widehat{EAD}_w^{det,\tau}$) through Eq. 6.21, where the superscript *det* replaces *if*.
- ✓ ***Step 6:*** Compute the economic effectiveness of the FEWS by using Eq. 6.1.
- ✓ ***Step 7:*** Build the contingency tables according to the FPC (Table 5.3) and estimate the reliability of the FEWS in terms of POD and FAR (Table 5.4).

6.6.3 Imperfect probabilistic forecast scenario

The PrFS assumes that the warning decision and the inundation forecasting is based on a warning level $\hat{y}_w$ estimated from $f(y|\hat{y})$. The algorithm used to estimate the metrics that define the reliability and effectiveness is detailed as follows.

- ✓ ***Step 1***: Uses the MCFG to sample n bivariate values from $f(y,\hat{y})$ and its associated conditional distribution $f(y|\hat{y})$ based on the BLND algorithm described in section 4.4.4.
- ✓ ***Step 2:*** Define a set of values between 0 and 1, e.g., [0.1,0.15,0.2 ,… 0.975] to define PT_k as the value assumed for the probabilistic threshold PT within this set.
- ✓ ***Step 3:*** Based on PT_k, define for each conditional distribution $f(y|\hat{y}_i)$ the warning level $\hat{y}_{w,i}$, and use the FWDC to simulate the warning decision according to the PDR (Eq. 5.3), i.e., warn only if $\hat{y}_{w,i} > y_T$.
- ✓ ***Step 4:*** Use the FTC described in Chapter 5 to define each pair $(y_i, \hat{y}_{w,i})$ as $(y_i^h, \hat{y}_{w,i}^h)$, $(y_i^m, \hat{y}_{w,i}^m)$, and $(y_i^f, \hat{y}_{w,i}^f)$, i.e., as hit, miss, and false alarm events, respectively (Table 5.2). In addition, define each pair $(y_i^h, \hat{y}_{w,i}^h)$ as $(y_i^{h1}, \hat{y}_{w,i}^{h1})$ and $(y_i^{h2}, \hat{y}_{w,i}^{h2})$, i.e., hits 1 and hits 2, respectively.
- ✓ ***Step 5:*** Use the RIC to define the economic consequences of each pair $(y_i^{h1}, \hat{y}_{w,i}^{h1})$, $(y_i^{h2}, \hat{y}_{w,i}^{h2})$, $(y_i^m, \hat{y}_{w,i}^m)$, and $(y_i^f, \hat{y}_{w,i}^f)$ according to Table 6.1
- ✓ ***Step 6:*** Estimate the ED in the PrFS associate with PT_k ($\widehat{ED}_{w,k}^{pr,\tau}$) through Eq. 6.21, where the superscript *pr* replaces *if*.
- ✓ ***Step 7:*** Repeat Steps from 3 to 6 for each PT_k.
- ✓ ***Step 8***: For each $ED_{w,k}^{pro,\tau}$, compute the economic effectiveness $E_{w,k}^{pr,\tau}$ of the flood warning system by using Eq. 6.1.
- ✓ ***Step 9***: Define PT^* as the PT_k associated with $max(E_{w,k}^{pr,\tau})$.
- ✓ ***Step 10:*** Estimate the reliability of the FEWS by repeating Step 7 of the DFS case above.

6.7 Sensitivity analysis

In the prior sections, the criteria, metrics, and algorithms considered in the hydro-economic EAD model to explore the effectiveness and reliability of a FEWS were described. In this last section of the chapter, the sensitivity of several parameters of the model that represents the main assumptions in the estimation of these flood warning performance attributes is analysed. This is done through several sensitivity experiments based on the one-at-a-time (OAT) method (Pianosi *et al.*, 2016), which, in essence, varies/perturbs the input parameters of the hydro-economic ED model from its reference parameter values (baseline scenario) one at a time and assess the impacts on the metrics used to define the reliability and economic effectiveness Thus, in this section, the parameters and the flood warning performance attributes that define the baseline scenario are first described, and then, the results of the sensitivity analysis are shown.

6.7.1 Baseline scenario

The hydro-economic ED model can be considered an extension of the framework used in Chapter 5 to explore the reliability of flood warning systems based on inundation forecasting. Thus, the hydro-economic ED model uses three additional parameters (γ,α, ϑ) to those considered in the framework shown in Chapter 5 (Table 5.6).

Table 6.2: Assumed input parameter values for the baseline scenario

Abbreviation	Description	Value adopted	Parameter associated with:
μ_y	Mean of y.	2.51	The river basin
σ_y^2	The coefficient of variation of y.	0.20	
y_o	Location parameter of y	2.03	
γ_y	The average number of peaks per year	1.60	
L	Basin lagtime L	6 hrs	
$\mu_{\hat{y}}$	Mean of the forecasts of y	μ_y	Flood forecasting and warning system
$\sigma_{\hat{y}}^2$	The variance of the forecasts of y	σ_y^2	
$\hat{y}_o$	Location parameter of the forecasts of y	y_o	
$\gamma_{\hat{y}}$	The average number of peaks per year of the forecasts of y	γ_y	
IP_ρ	The inflection point of the lead time-performance function	0.85	
τ	Lead time	6 hr	
γ	A parameter that defines the cost of the warning response as a percentage of potential benefits.	0.1	
T_{SoP}	Return period associated with the flooding threshold y_T.	5 years	The at-risk community
$n_{houses}^{at-risk}$	Total number of houses at risk in the benefit area	1000	
ϑ	The monetary value of residential contents	1 unit.	
α	Efficiency parameter of the proactive action	0.5	
n	Sample size or number of simulated events	10000	Monte Carlo simulation

Table 6.2 depicts the input parameters of the model and the values which define the baseline scenario. They are split according to the component they represent. The baseline hydro-economic ED model assumes that the cost of the warning response C_w^τ represents 10% of the value of the damage avoided, i.e., γ =0.1 in Eq.6.17, and the efficiency of the proactive action α is assumed to be 50 %. Thus, the values set for γ and α were 0.1 and 0.5, respectively. The monetary value for the residential contents in each floodplain property was assumed to be 1 unit. The adopted values

for the remaining parameters, which match with the framework shown in Chapter 5, were the same as those considered in Table 5.6. The rationality of these adopted values is explained in section 5.7.1, and it is also used in this work to represent the baseline scenario in this sensitivity analysis. Finally, the number of simulated events was set to 500,000 for the Monte Carlo estimates. It is worth noting that Figure 5.7 can also be used to represent the effect of the standard of protection SoP on functions that describe the flood damage without the warning services (Eq. 6.16) and the net damage in the PFS (Eq. 6.17). In this case, the impact represents the flood damage or net flood damage, and the SoP shifts the value at which the flood impact begins to the value of T_{SoP}. The results of the baseline scenario are explained as follows.

The analysis of the baseline scenario, under the criteria mentioned above, is summarised in Figure 6.6 and Figure 6.7, whose main values are shown in Table 6.3. Figure 6.6 shows the results of the optimization procedure used to find the optimal probabilistic threshold PT^* for the probabilistic decision rule PDR in the PrFS based on the algorithm described in section 6.6.3. The solid grey curve indicates how the value of E_w changes when considering several values of the probabilistic threshold PT for the PDR. The figure also shows the values of E_w based on the PFS and DFS obtained from the algorithms described in sections 6.6.1 and 6.6.2, represented by a solid black and blue line, respectively. Figure 6.7 shows, on the other hand, the reliability of the flood warning system in terms of POD and FAR based on the FPC for the DFS and PrFS. These figures can be used to analyse how the effectiveness and reliability are related.

Table 6.3: Flood warning performance attributes of the baseline scenario
These values correspond to the Figures 6.6 and 6.7

Scenario	Decision rule	PT [-]	Lead time [hr]	POD [-]	FAR [-]	E_w [%]
DFS	DFDR	-	6	0.60	0.39	4.13
PrFS	PDR	0.54		0.84	0.61	5.24
PFS	-	-		1	0	7.5

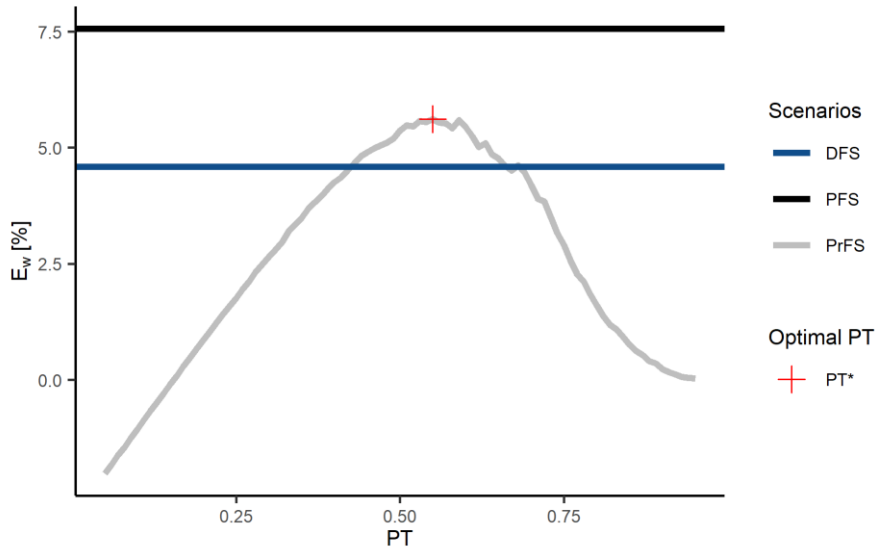


Figure 6.6: Estimation of the economic effectiveness of the baseline model under different forecasting scenarios.

The grey line shows the results of the optimization procedure used to find the optimal value of the probabilistic threshold PT used in the warning strategy of PrFS; the optimal value PT^* is described with a red cross. The effectiveness of the DFS and PFS are illustrated with horizontal lines. *DFS: Deterministic forecast scenario; PFS: Perfect forecast scenario; PrFS: Probabilistic forecast scenario; PT: probabilistic threshold.*

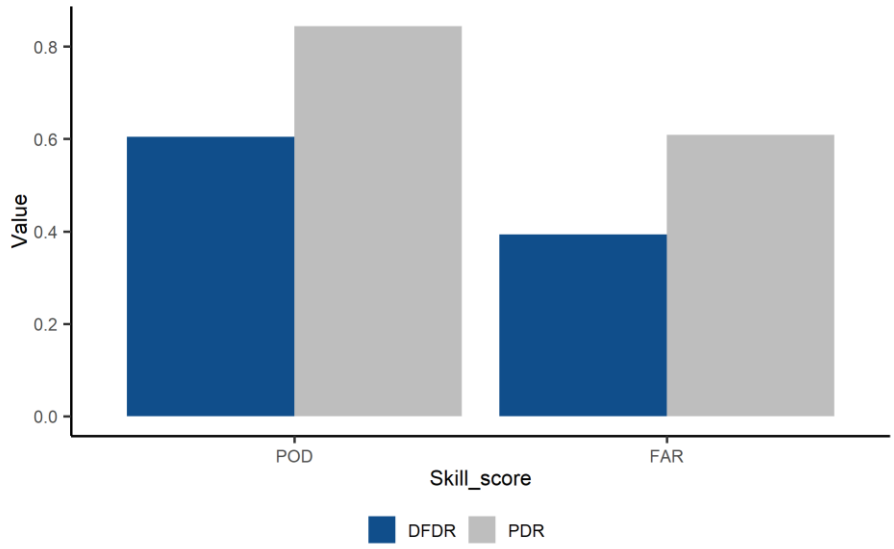


Figure 6.7: Estimation of the reliability of the baseline model under the imperfect forecasting scenarios

The reliability is evaluated in terms of POD and FAR and evaluated based on the FPC (5.4.2). *DFDR: Decision rule used in the deterministic forecast scenario; PDR: Decision rule used in the probabilistic forecast scenario; POD: probability of detection; FAR: false alarm ratio*.

The PFS delivers the maximum economic benefits one can obtain from the FEWS. Figure 6.6 shows that, as expected, the value of E_w for this scenario is highest. The economic effectiveness of the PrFS is higher than that of the DFS; it is due to the decision rule used in that scenario, i.e., the PDR, assumes that the Warner acts to increase the economic benefits, i.e., E_w, in the at-risk community. In terms of

reliability, Figure 6.7 shows that this flood warning performance attribute of the baseline model based on the FPC for the DFS is in the Low category (section 5.7.2.1), as the values of POD and FAR are higher and lower than 0.5, respectively. The probabilistic decision rule PDR, on the other hand, delivers a FAR value higher than 0.5, which might be considered not useful in operational FEWS. These results show that when the cost of the warning response $C_w^\tau(\hat{d})$ is low, represented in this case as 10% of the forecast economic benefits, i.e., γ =0.1, one can "sacrifice" the reliability of the FEWS in terms of increasing false alarms, to increase the economic effectiveness, as avoiding misses has the biggest impact on damage mitigation and E_w.

An interesting point here is to compare the reliability results of the PDR (0.84 and 0.61 for POD and FAR, respectively (Table 6.3)) based on the optimisation criterion adopted in this Chapter, i.e., maximizing E_w, with the ones obtained in Chapter 5 based on the POD-FAR optimisation criterion (0.79 and 0.55 for POD and FAR, respectively (Table 5.7)). The results indicate that both optimization criteria generate high values of POD and FAR and that the values of PDR based on the optimization of E_w are greater than those obtained in terms of POD-FAR. As was concluded in Chapter 5, the high values of FAR associated with the POD-FAR criterion are due to this optimization criterion does not place any restriction on the value of FAR, and it looks for the biggest difference between POD and FAR, regardless of the value of the latter. On the other hand, as was concluded above, the high FAR value associated with the optimization criterion adopted in this Chapter is due to the low cost of the warning response $C_w^\tau(\hat{d})$ (γ =0.1). Note, however, that the reliability of flood warnings associated with this criterion is strongly controlled by $C_w^\tau(\hat{d})$; therefore, one would expect that when increasing $C_w^\tau(\hat{d})$, the values shown in Table 6.3 will change; it will be analysed in section 6.7.2.3.

As was discussed in Chapter 5 (section 5.7.2.1), the effect of having high values of FAR in a FEWS is known as 'cry wolf', which has to do with the disregarding of flood warnings due to their loss of credibility as a result of the high percentage of false alarms. However, Barnes *et al.* (2007) advocate that there is little evidence that a high value of FAR causes users to ignore warnings of severe events.

6.7.2 Sensitivity experiments

In this last sub-section of the chapter, the sensitivity of several parameters with respect to the results shown in Table 6.3 is explored. This was done based on three sensitivity experiments.

6.7.2.1 Sensitivity experiment 1

In principle, there is a trade-off between the economic benefits and reliability of a FEWS when analysing these two flood warning performance attributes as a function of the lead time, i.e., the longer the lead time, the higher the damage saved and the lower the reliability, respectively (Schröter *et al.*, 2008). However, it has been demonstrated that when including the net damage associated with forecast uncertainty (section 6.2), that trade-off can be undermined. For example, Verkade and Werner (2011) showed that the expected annual benefits of a FEWS could decrease when the lead time and reliability are long and low, respectively, due to the economic consequences of false alarms and misses. This experiment aims to explore the relationship between these two flood warning performance attributes for each forecasting scenario. For this purpose, it is assumed that the forecasts of several FEWSs with several lead times shorter or longer than that of the baseline model (6hr) (Table 6.2) are represented by the same moments of $\hat{y}$ (which are taken to be equal to those of y), and that the forecast uncertainty is only controlled by $\rho_{y\hat{y}}$. In this experiment, $\rho_{y\hat{y}}$, defined by the lead time-correlation function in Eq. 4.57, describes the forecast uncertainty relationship with lead time which, at the same time, reduces or increases the forecast uncertainty, respectively. Table 6.4 gives a description of this experiment and the range of values of τ to be analysed.

Table 6.4: Description of sensitivity experiment 1

Parameter to be modified	Lower bound	Baseline	Upper bound	Aim
τ	3 hr (-50%)	6 hr	24 hr (+200%)	Analyse how the flood warning performance attributes (reliability and economic effectiveness) of the baseline model, for each forecasting scenario, changes as the lead time changes.

The results of this experiment are shown in Figure 6.8 and Figure 6.9, which shows how these flood warning performance attributes of the baseline scenario change as the lead time τ changes. The results indicate that the economic benefits of a FEWS based on a PFS increase as τ increases, and that, when considering the net damage associated with forecast uncertainty, there is not a trade-off between the reliability and economic benefits; they show that for the DFS and PrFS there is an optimal lead time (τ = 12hrs) that maximizes the economic benefits in the at-risk community. Furthermore, the DFS-based and PrFs-based economic benefits of FEWSs with short lead times (τ values shorter than the basin lagtime L) are close to the maximum economic benefits that can be achieved by the system, i.e., those obtained from the PFS. The rationality of all these results is explained as follows.

As was explained in section 6.3.1, the PFS-based economic effectiveness represents the maximum economic benefits that can be achieved by the FEWS as they are not affected by the net damage associate with forecast uncertainty (Figure 6.2c). That means this scenario delivers the highest values of E_w (solid black line) and represents the theoretical limit of the economic benefits, which according to Figure 6.8 ranges from 4 to 15% approximately for FEWS with lead times between 1 and 24h. Note that these values depend on the adopted values for the economic parameters γ and α in the baseline FEWS (Table 6.2). Note also that the net damage associated with forecast uncertainty is present in false alarms, misses, *and* hits (Figure 6.2), and they increase ED_w in Eq. 6.1, which, in turn, decrease the economic effectiveness E_w. Since the DFS and PrFS are characterised by the presence of these events, neither of these scenarios can overcome the economic benefits of the PFS.

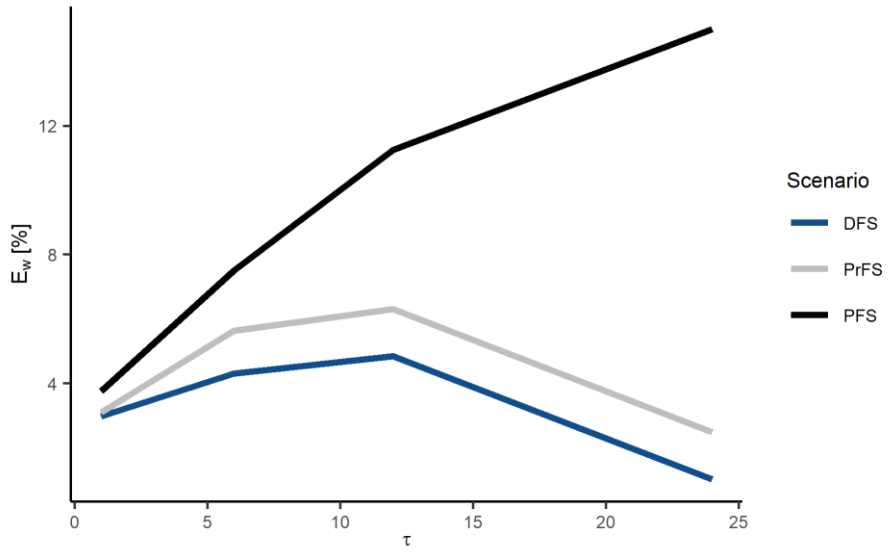


Figure 6.8: Analysis of the effectiveness of the baseline model based on the sensitivity experiment 1

In this experiment, the effectiveness of the baseline model is analysed by reducing or extending the lead time τ, which, at the same time, reduces or increases the forecast uncertainty, respectively. *DFS: Deterministic forecast scenario; PFS: Perfect forecast scenario; PrFS: Probabilistic forecast scenario*.

According to the results shown in Figure 6.8, the economic effectiveness of a deterministic FEWS (solid blue line) with forecasting lead times between 1 and 24h can range from 1% and 4% approximately. The economic effectiveness increases up to 12 hours but then decreases. These results also indicate that the net damage associated with forecast uncertainty increases as the lead time τ increases beyond 12 hours since these economic benefits move away from the PFS-based benefits and drop as the lead time τ increases. Figure 6.9 also shows that the reliability of flood warnings based on deterministic forecasts decreases as the lead time τ increases, which is expressed with lower and higher values of POD and FAR, respectively. There is not, therefore, a trade-off between the economic benefits and reliability for the DFS case, since the former decreases after a certain lead time. Thus, the DFS-based effectiveness allows an optimal lead time to be analysed. Figure 6.8 shows that this lead time is 12 hours approximately.

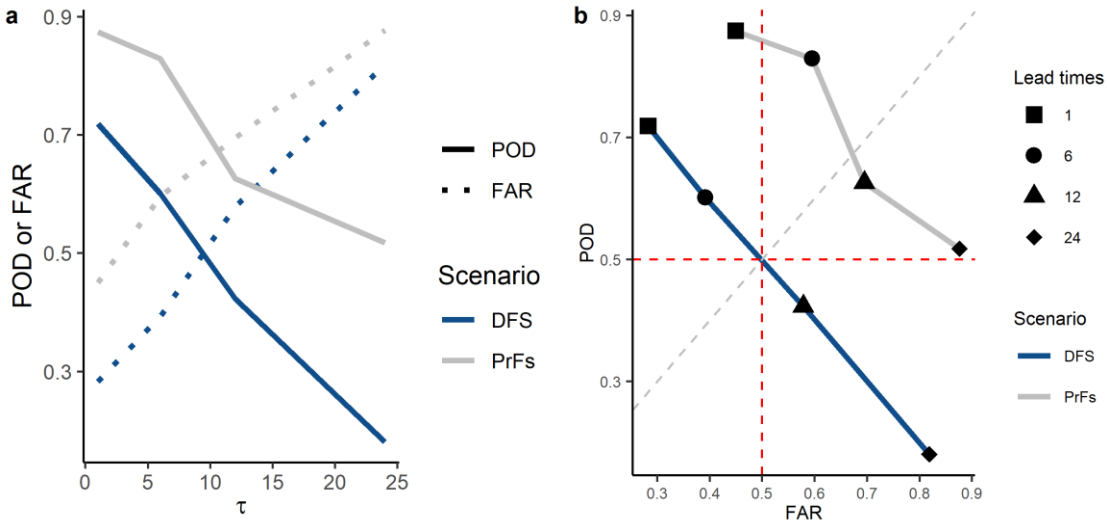


Figure 6.9: Analysis of the reliability of the baseline model based on the sensitivity experiment 1

In this experiment, the reliability of the baseline model is analysed by reducing or extending the lead time τ which, at the same time, reduces or increases the forecast uncertainty, respectively. The reliability is evaluated in terms of POD and FAR based on the FPC (5.4.2). Figure a shows FAR and POD as a function of the lead time τ. The values of this figure are summarised in the FAR-POD curves shown in b.

DFS: Deterministic forecast scenario; PFS: Perfect forecast scenario; PrFS: Probabilistic forecast scenario; POD: Probability of detection; FAR: False alarm ratio; FPC: floodplain property criterion.

The economic effectiveness of the DFS with short lead times (τ values lower than the assumed basin lagtime L (6hr)) are close to the PFS-based benefits because their forecast uncertainty is small, and, therefore, the (negative) impact of the net damage associated with forecast uncertainty (Figure 6.2) on the economic effectiveness is small. This small forecast uncertainty also produces high reliability (Figure 6.9a). The hydro-economic ED model assumes that these systems have small forecast uncertainty because they use forecasting models based on observed data (real-time data), such as gauge-based quantitative precipitation (QPE), and may also employ forecast updating. The forecast uncertainty of FEWSs with long lead times (τ values beyond L) drastically increases because the forecasting models have to be forced with quantitative precipitation forecasts (QPFs). The hydro-economic ED model represents this by decreasing the value of $\rho_{y\hat{y}}$ according to the lead time-correlation function shown in Figure 4.14 (note that in this experiment, the moments of $\hat{y}$ are assumed to be the same for each lead time τ, and the forecast uncertainty is only controlled by $\rho_{y\hat{y}}$).

The considerable forecast uncertainty of deterministic FEWSs with long lead times reflect the difference between y and $\hat{y}$, (the forecast error) and therefore, between n_{houses}^{warned} and $n_{houses}^{flooded}$, reducing the benefits of 'good' decisions (hit events) and

increasing the presence of 'wrong' decisions (false alarms and misses events) in the system. That increases the net damage associated with forecast uncertainty (Figure 6.2) and, in turn, decreases the effectiveness E_w. The economic benefits of these systems can be lower than those obtained from FEWSs with shorter lead times. A decrease in E_w means that the value of ED_w approaches that of ED_{nw} (Eq. 6.1); if the former is higher than the latter, the forecasts of the FEWS do not have value in the at-risk community. Note that, even though the DFS-based results are affected by considerable forecast uncertainty for long lead times, which produces low reliability (Figure 6.9), these systems still have value in the at-risk community since the value of E_w is higher than zero.

Like the DFS-based results, the PrFS-based economic effectiveness (solid grey line in Figure 6.8) shows an optimal lead time of approximately 12 hours. These results indicate that an imperfect FEWS is characterised by a lead time that represents the balance between an adequate time to act and a reasonably good forecast. Likewise, the economic effectiveness of probabilistic FEWSs with short lead times approaches the PFS-based effectiveness due to the small forecast uncertainty but is slightly better than that obtained from the deterministic FEWS. However, as the lead time increase, the economic effectiveness of the PrFS separates from deterministic-forecast-based economic effectiveness as is clearly better. These results tell us that the benefits of using probabilistic information in a FEWS in comparison to the deterministic information are most noticeable in FEWSs based on forecasts with relatively high forecast uncertainty.

Figure 6.8 shows that these economic benefits of the PrFS can range from 3% to 5.5 % approximately. That increase (1.5% with respect to the maximum deterministic-forecast-based value) is due to the adopted warning strategy in this scenario, which assumes that the warner acts to increase the economic benefits in the at-risk community (section 6.4.2). To increase the economic effectiveness of FEWSs with longer lead times, the Warner has to warn more frequently with respect to the DFS. This is intended to avoid the high economic impact of having flooded houses not being warned in the at-risk community, characterised as 'misses' in the FPC (section 5.4.2). That strategy, however, increases the probability of having warned houses that are not subsequently flooded (section 5.4.2). These probabilities are described by FAR, and that is the reason why Figure 6.9 shows high values of

these metrics. That means that due to the low cost of the warning response $C_w^{\tau}(\hat{d})$ of the baseline scenario (defined by the parameter r=0.1 (Eq. 6.13)), the best economic strategy is to warn more frequently accepting or "sacrificing" the low economic impact of having warned houses that are not subsequently flooded to avoid the high economic impact of having flooded houses not being warned. Note that this warning strategy also reduces the probability of having flooded houses that were not warned, characterised as missed events in the FPC(section 5.4.2). These probabilities are described in Figure 6.9 by 1-POD (grey line). The values of this probability indicate that even though the warning strategy of the PrFS warns more frequently for long τ values with respect to the DFS, the probability of having flooded houses that are not warned is still considerable, reflecting the high uncertainty.

One should bear in mind that, even though the warning strategy in the PrFS produces greater economic benefits than the DFS-based results, the high values of FAR might produce a loss of credibility in the FEWS ('cry wolf' effect), and an increase in worry for flood plain residents (psychological impact). That might decrease the effectiveness of the proactive action, which is represented in the hydro-economic ED model by the parameter α. The adopted value of this parameter for the baseline scenario is 0.5 (Table 6.2), which is a fixed value in this experiment for each lead time τ, regardless of the value of FAR. The PrFS-based results do not take this into account, and the economic effectiveness depicted in Figure 6.8 (solid grey line) might be, thus, overestimated. However, as was mentioned above, Barnes *et al.* (2007) advocate that there is little evidence that a high value of FAR causes users to ignore warnings of severe flood events.

6.7.2.2 Sensitivity experiment 2

The sensitivity experiment 1 explores the economic effectiveness and reliability of the baseline FEWS based on forecasts for several τ values. That experiment assumes that these forecasts are represented by the same moments, and, therefore, the forecast uncertainty can be only modified through the value of $\rho_{y\hat{y}}$, which, in turn, is related to τ through the lead time-correlation function shown in Figure 4.14. One can also use this function to evaluate the impact of decreasing or increasing the forecast uncertainty on the results shown in the sensitivity experiment 1. That is done by considering the baseline parameters (Table 6.2) and modifying the

parameter IP_ρ , where the latter represents the inflexion point of the lead time-correlation function, and exploring the effect as a function of τ. A higher or lower value of this parameter means a smaller or larger forecast uncertainty (better or worse forecast quality) for each lead time τ, respectively. Thus, in this experiment, the sensitivity of IP_ρ with respect to the reliability and economic effectiveness of the baseline FEWS characterised by different τ values, under different forecasting scenarios, is analysed. Since research works indicate that an improvement of 10% per decade in the forecast performance is achievable in FEWSs (Pappenberger *et al.*, 2015), this parameter was modified by a percentage of ± 10%. Table 6.5 gives a description of this experiment and the range of values to be explored. In this case, several values within the range τ were considered, whereas for the IP_ρ, only the values of the lower and upper bound were included in the analysis.

Table 6.5: Description of sensitivity experiment 2

Parameter to be modified	Lower bound	Baseline	Upper bound	Aim
τ	3 hr (-50%)	6 hr	24 hr (+200%)	Analyse how the increase or reduction of forecast uncertainty impact on the flood warning performance attributes (reliability and effectiveness) of a FEWS with a given lead time.
IP_ρ	0.77 (-10%)	0.85	.94 (+10%)	

The results of the experiment are depicted in Figure 6.10 and Figure 6.11, which shows how the economic effectiveness and reliability of the baseline FEWS (τ= 6hr) based on forecasts for several τ values change when modifying the values of IP_ρ. The dashed blue and grey lines in Figure 6.10 indicate how the DFS-based and PrFS-based effectiveness, respectively, change when modifying the baseline value of IP_ρ (.85) to 0.94 (Figure 6.10Figure 6.10) and 0.77 (Figure 6.10b). Similarly, the solid green and grey lines in Figure 6.11 indicate how POD and FAR change when making these changes to the baseline value of IP_ρ. Note that in this experiment, the moments of $\hat{y}$ are assumed to be the same for each lead time τ, and are equal to those of y, so the forecast uncertainty is only controlled by $\rho_{y\hat{y}}$.

Figure 6.10 shows that the sensitivity of the DFS-based effectiveness with respect to IP_ρ is more or less the same (variation of approximately ±1.5% of the baseline value) for all lead times, whereas the PrFS-based effectiveness shows poor sensitivity to the lower bound IP_ρ values and high sensitivity to the upper bound IP_ρ

values for long lead times τ (τ >12 hr). These results indicate the benefits of improving the forecast performance (expressed here as an increase of 10% of the IP_ρ value), in economic terms, are most noticeable in probabilistic FEWSs whose forecast uncertainty is relatively high, expressed in this experiment as long τ values.

In terms of reliability, Figure 6.11 shows that the DFS-based reliability's sensitivity is higher for short lead times (τ < 9 hr); the same occurs for the PrFS-based reliability, particularly for FAR values. This suggests that the benefits of improving the forecast performance (expressed here as an increase of 10% of the IP_ρ value), in terms of reliability, are most noticeable in (deterministic and probabilistic) FEWSs whose forecast uncertainty is relatively low, expressed in this experiment as short τ values.

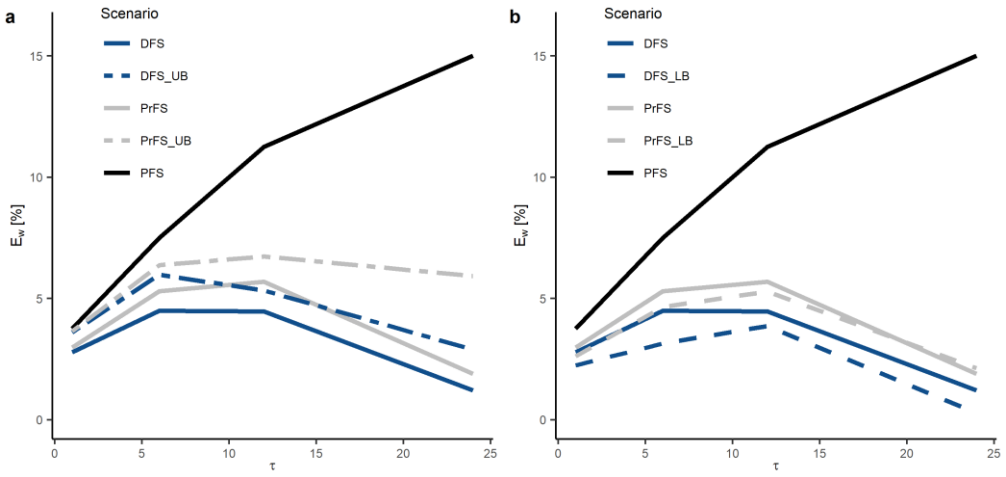


Figure 6.10: Analysis of the economic effectiveness of the baseline model based on the sensitivity experiment 2

In this experiment, the impact of the forecast uncertainty, controlled by correlation through IP_ρ on the effectiveness of the baseline model is analysed. The dashed blue and grey lines indicate when the forecast uncertainty is improved (a) or deteriorated (b), respectively, in the DFS and PrFS. In this experiment, the forecast uncertainty associated with a given lead time τ is changed by modifying the value of the parameter IP_ρ.

DFS: Deterministic forecast scenario; PFS: Perfect forecast scenario; PrFS: Probabilistic forecast scenario; DFS_UB: Deterministic forecast scenario based on the upper bound of IP_ρ; PrFS_UB: Probabilistic forecast scenario based on the upper bound of IP_ρ; DFS_LB: Deterministic forecast scenario based on the lower bound of IP_ρ; PrFS_UB: Probabilistic forecast scenario based on the lower bound of IP_ρ

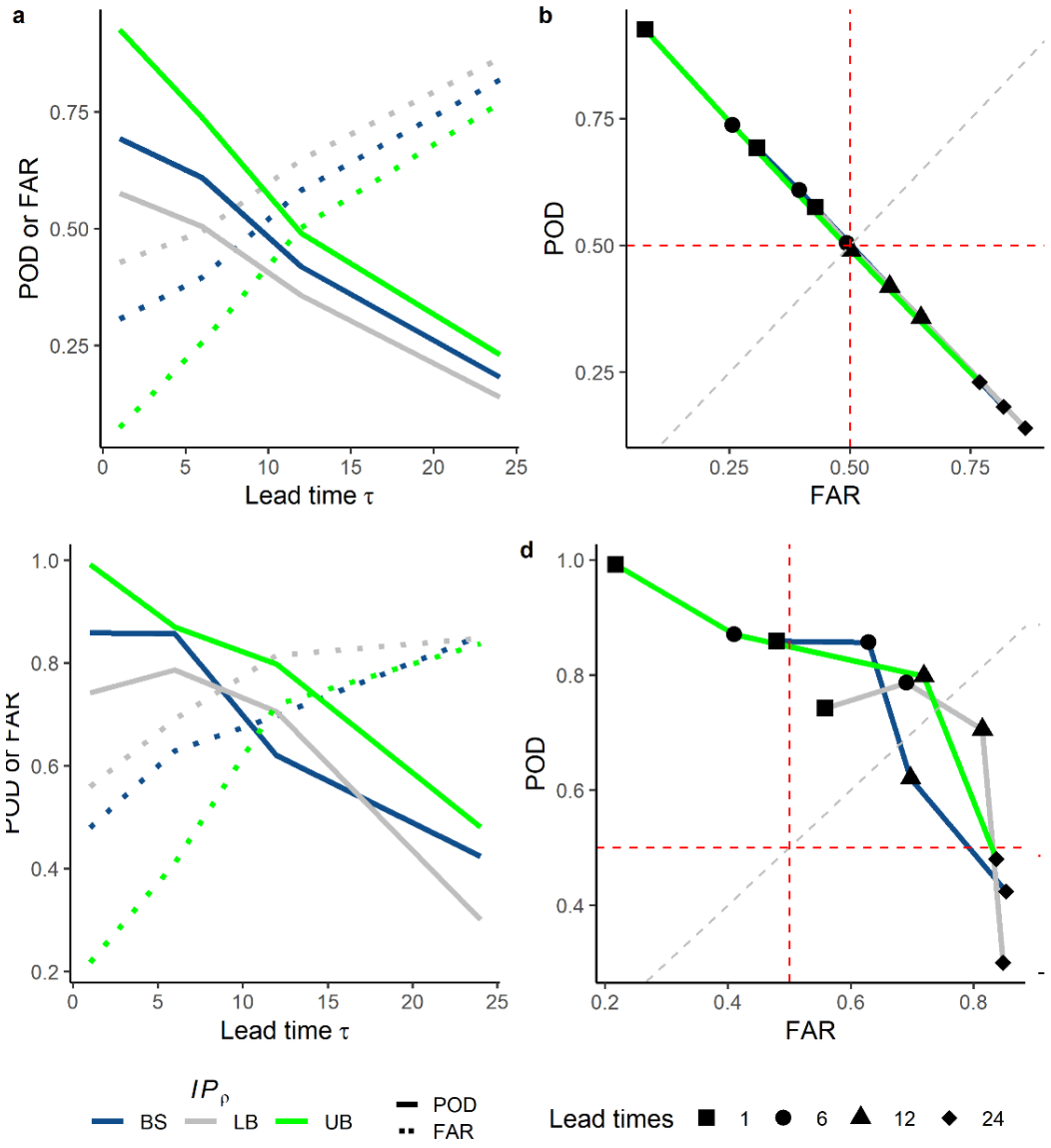


Figure 6.11: Analysis of the reliability of the baseline model based on the sensitivity experiment 2

In this experiment, the impact of the forecast uncertainty on the reliability of the baseline FEWS based on several lead times is analysed. The reliability is expressed in terms of POD and FAR based on the FPC. a) DFS-based results evaluated based on the FPC as a function of the lead time τ; these values are summarised in the POD-FAR curve shown in b). The same figures are used to show the PrFS-based results (c and d).

LB: lower bound of IP_ρ ; BS: Baseline scenario; UB: upper bound of IP_ρ; POD: Probability of detection; FAR: False alarm ratio

6.7.2.3 Sensitivity experiment 3

In this last experiment, the two input parameters that control the economic effectiveness, i.e., γ and α, are perturbed. The parameter γ controls the cost of the warning response $C_w^\tau(\hat{d})$ and α represents the performance of the proactive action. The aim of the third sensitivity experiment is, therefore, to analyse the sensitivity of the economic effectiveness to perturbations of these economic parameters.

It is estimated that the emergency costs represent 10% of the property damage(Penning-Rowsell *et al.*, 2020); therefore, values between 0.05 and 0.5 for γ were considered to analyse how an increase or decrease in $C_w(\hat{d})$ affects the economic effectiveness E_w and the reliability of the PrFS-based results. On the other hand, the parameter α was varied between 0 and 1 to explore how an increase or decrease in the performance of the proactive action impacts on E_w. Table 6.6 gives a general description of this experiment.

Table 6.6: Description of sensitivity experiment 3

Parameter to be modified	Lower bound	Baseline	Upper bound	Aim
γ	0.05	0.1	0.5	Analyse the sensitivity of the economic parameters of the hydro-economic ED model when estimating the economic effectiveness.
α	0.1	0.5	1	

Parameter α

Figure 6.12 shows the results of the sensitivity experiment in terms of economic effectiveness E_w, and Figure 6.13 in terms of reliability for the economic parameter α reflecting the efficiency of the proactive action. Figure 6.12 shows, as expected, that the economic effectiveness E_w increases as α increases. A perfect scenario of the baseline FEWS (perfect response and forecast) produces economic effectiveness of 15% approximately. Based on the Carsell functions (Figure 6.5), this represents the maximum economic benefits that can be achieved by a FEWS with a lead time of 6 hours. This economic effectiveness E_w *cannot* be achieved by imperfect FEWSs due to the net damage associated with forecast uncertainty and the inefficiency of the response. Figure 6.12 shows that if the baseline FEWS improves the efficiency of the proactive action to 70%, it can increase the economic effectiveness E_w to 5.5 and 7.5 % approximately for the DFS and PrFS, respectively. This figure also shows that the economic effectiveness for the DFS and PrFS are practically the same for low values of α (α<0.25). These results indicate the benefits of using probabilistic information in a FEWS with respect to the deterministic information are most noticeable in FEWSs where the potential economic benefits of the system are relatively high.

Figure 6.13 shows that the reliability results for both the DFS and PrFS cases are not sensitive to perturbations to the parameter α. Note how, in the DFS, for all

values of α, the POD and FAR values are the same as for the baseline FEWS, 0.61 and 0.39 (Table 6.3), respectively. This is represented by a horizontal line in Figure 6.13a and by several points clustering at the point representing the baseline results in the POD-FAR curve. For the case of the PrFS, even though the sample size used is large (500,000), there is some instability in the results. However, one can note that the values of POD and FAR delivered by the probabilistic decision rule PDR are on average 0.85 and 0.6, respectively.

Based on these results shown in Figure 6.12 and Figure 6.13, one can also conclude that the performance of the proactive action is the main factor controlling the economic benefits of the FEWS. Note that, whatever the values of POD and FAR (this includes the baseline values and good values of POD and FAR), if the performance of the proactive action is poor, the economic effectiveness of the FEWS will be low.

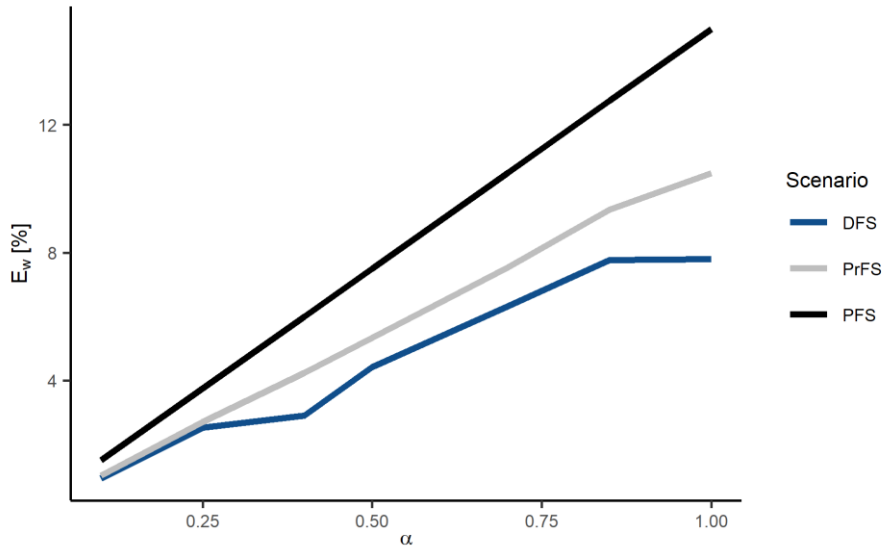


Figure 6.12: Sensitivity analysis of the performance of the proactive action on the economic effectiveness of a FEWS

This figure shows how the economic effectiveness of the baseline FEWS changes as the parameter α changes for several forecasting scenarios.

DFS: Deterministic forecast scenario; PFS: Perfect forecast scenario; PrFS: Probabilistic forecast scenario.

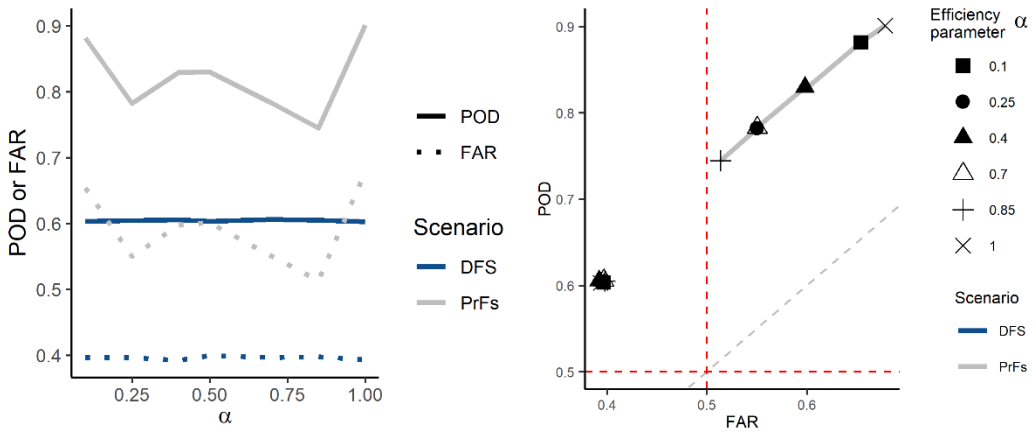


Figure 6.13: Sensitivity analysis of the performance of the proactive action on the reliability of the baseline FEWS

This figure shows how the reliability of the baseline FEWS changes as the parameter α changes for several forecasting scenarios. The reliability is expressed in terms of POD and FAR. a) Results evaluated based on the FPC as a function of the lead time τ; these values are summarised in the POD-FAR curve shown in b).

DFS: Deterministic forecast scenario; PFS: Perfect forecast scenario; PrFS: Probabilistic forecast scenario; POD: Probability of detection; FAR: False alarm ratio

Parameter γ

Figure 6.14 shows the results of the sensitivity experiment in terms of economic effectiveness E_w, and Figure 6.15 in terms of reliability for the parameter γ. This parameter controls the cost of the warning response $C_w^\tau(\hat{d})$, and represents the value adopted by the hydro-economic ED model to estimate $C_w^\tau(\hat{d})$ as a percentage of the economic benefits of the FEWS if a forecast with the magnitude $\hat{y}$ or $\hat{y}_w$ occurs in the DFS or PrFS, respectively (section 6.4.3.5). Figure 6.14 shows that, as expected, the higher the value of γ, the lower E_w and vice versa. This occurs because, when increasing or decreasing $C_w^\tau(\hat{d})$, the economic consequences of false alarms and hits increases or decreases, respectively. Note that, even for the high upper bound value of γ considered (50%), E_w is still positive (> zero) for both scenarios. Note, however, that with the increase of γ, the PrFS does not impact E_w in the same proportion as it impacts in the DFS. It is due to the economic warning strategy adopted in the PrFS which modifies the warning criterion to avoid E_w decreasing significantly. That is why reliability is sensitive to γ for the PrFS. Figure 6.15 shows that when increasing $C_w^\tau(\hat{d})$, the best economic strategy is to warn less frequently as the values of POD and FAR decrease. If $C_w^\tau(\hat{d})$ decreases, the opposite effect occurs. These results tell us the importance of estimating $C_w^\tau(\hat{d})$ when quantifying the economic benefits of a FEWS, as an optimal warning strategy depends on this variate. The sensitivity of E_w to the perturbations in γ is nil for the DFS which is

represented by a horizontal line in Figure 6.15a, and by several points that cluster at the point representing the baseline results in the POD-FAR curve (Figure 6.15b). For the case of the PrFS, even though the sample size used is large (500,000), there is some instability in the results.

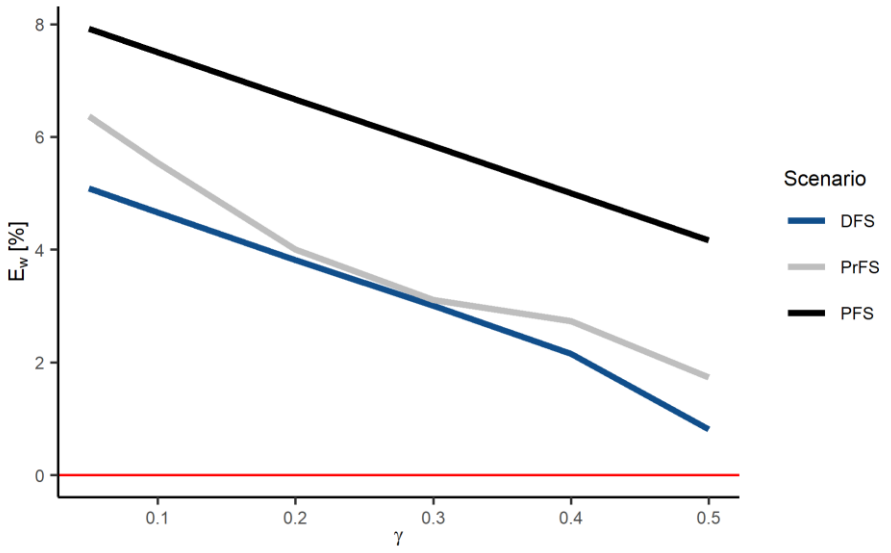


Figure 6.14: Sensitivity analysis of the parameter γ on the economic effectiveness of the baseline FEWS

In this experiment, the impact of the parameter γ on the economic effectiveness of the baseline model is analysed for all warning scenarios.

PFS: Perfect forecast scenario; DFS: Deterministic forecast scenario; PrFS: Probabilistic forecast scenario

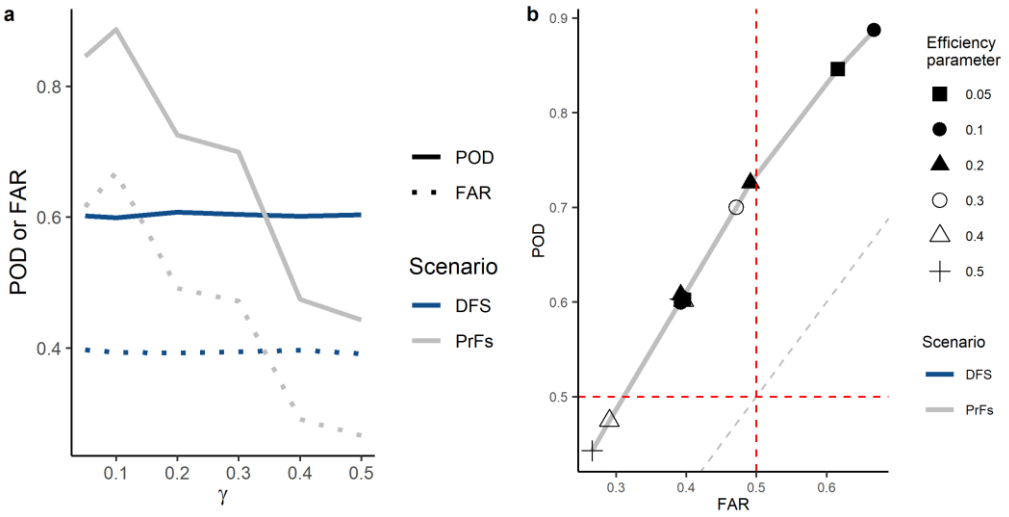


Figure 6.15: Sensitivity analysis of the reliability of the baseline FEWS to changes in the parameter γ.

In this experiment, the impact of the parameter γ on the flood warning reliability of the baseline model is analysed for the imperfect warning scenarios.

DFS: Deterministic forecast scenario; PrFS: Probabilistic forecast scenario; POD: Probability of detection; FAR: False alarm ratio.

6.8 Main findings

Thanks to the integrated framework's versatility, the economic consequences of forecast uncertainty in a wide range of possible situations could be included in the analysis. The results showed that the warning strategy is an important factor influencing the economic effectiveness E_w of an imperfect FEWS. This research showed that an optimal warning strategy based on probabilistic forecasts and an estimation of E_w produced by the FEWS for each potential flood can obtain greater benefits than those obtained with a deterministic-forecast warning strategy. Furthermore, it was observed that the cost of the warning response $C_w^\tau(\hat{d})$ controls the flood warnings' reliability and the warning strategy in a probabilistic FEWS where warning decisions are based on a probabilistic threshold. Depending on the value of $C_w^\tau(\hat{d})$, an optimal warning strategy for the system could be warning less or more frequently. If $C_w^\tau(\hat{d})$ is low, an optimal economic strategy could have high FAR values. Furthermore, this Chapter shows that a probabilistic-forecast-based-optimal warning strategy, in economic terms, should be set for each lead time τ since the forecast uncertainty and its associated economic consequences increase as τ increases. Furthermore, the results of this Chapter indicate the benefits of using a probabilistic warning strategy with respect to a deterministic one are most noticeable in FEWSs based on forecasts with relatively high forecast uncertainty and where the potential economic benefits of the system are relatively high.

The forecasting lead time τ was also an important factor influencing the economic effectiveness E_w of a (deterministic and probabilistic) FEWS. The results showed that, when increasing τ, the economic effectiveness E_w of an imperfect FEWS move away from the maximum economic effectiveness one can obtain from the system, i.e., those obtained from a perfect forecast scenario, due to the (negative) impact of the net damage associated with forecast uncertainty on the economic benefits. This research shows that an imperfect FEWS is characterised by an optimal lead time that represents the balance between an adequate time to act and a reasonably good forecast.

This Chapter also showed that the economic effectiveness E_w of a FEWS, based on the residents moving/ evacuating contents and a 6-hour lead time, can reach 15% in a perfect warning scenario (perfect forecast and response). This value cannot be achieved by imperfect FEWS due to the economic consequences of forecast

uncertainty and inefficiency of the proactive action. It was found that if one considers an efficiency parameter α of 70% and low cost of the warning response (10% of the forecast economic benefits), the economic effectiveness E_w of an imperfect FEWS, based on a 6h lead time, can reach 5.5 and 7.5 % for the deterministic and probabilistic forecasts, respectively. In this sense, it was concluded that the performance of the proactive action is an important factor influencing the economic effectiveness of a FEWS. Further, it was found that a FEWS could have good flood warning reliability but low economic effectiveness E_w due to the bad performance of the proactive action.

Finally, it was analysed that if an efficiency parameter α of 50% in the proactive action and a low cost of the warning response $C_w^{\tau}(\hat{d})$, with $\gamma = 0.1$, are considered in the FEWS, the economic effectiveness E_w of deterministic forecasts can range from 1 to 4% for lead times of 1 and 24 hours. These E_w values were improved by probabilistic FEWS whose E_w ranged from 3% to 5.5 % (section 6.7.2.1).

Chapter 7. The Nanjing Case Study

7.1 Introduction

The city of Nanjing has been chosen for the case study. Nanjing has suffered from severe pluvial flooding in recent years due to intense summer rain storms, notably in 2016, resulting in the inundation of a number of areas in the city. Most of these areas are polders that lie below the levels of the adjacent inner rivers, which are connected to outer rivers and ultimately the Yangtse river. Pumping systems are operated during storm events to remove water from the inner rivers to the outer rivers, to enable water to drain from the polder areas into the inner rivers. Even though storm warnings are issued in Nanjing, the pumping operations of these polders can be considered as reactive pump operations because they are mainly driven by the observed inflow. This case study is attractive because (a) rainfall forecasts can be used to conduct proactive pumping operations, and (b) the pumping represents a Risk Response Action (RRA) within the generic framework developed in this thesis. It, therefore, provides an excellent opportunity to demonstrate the value of probabilistic forecasts in making better decisions about pumping operations.

In this context, a framework to analyse different pumping strategies in the polder systems under different forecast scenarios based on 24-h forecasts has been developed. The simulation of the pumping operations is done through continuous simulation. Pumping strategies are assessed in terms of the average pumping cost and average inundated area of the rainiest month in Nanjing (July). Figure 7.1 illustrates the framework. It is made up of three components: i) a rainstorm-and-forecast generator (RFG) which generates hourly rainfall and deterministic and probabilistic 24-forecasts of runoff, ii) a flood warning decision component (FWDC) that simulates warning decisions based on the type of forecast generated by the RFG, and ii) a response and impact component (RIC) which represents forecast-based pumping strategies and gives an estimation of its impact on the polder system. Even though the algorithm designed to represent the response component can be considered simple, where the simulation time for a month can be done in a matter of seconds, the stochastic component of the framework, and the approach used to represent the RFG, requires a high computational effort. Most of

the analysis shown in this chapter has been done by using a computing cluster with 50 cores.

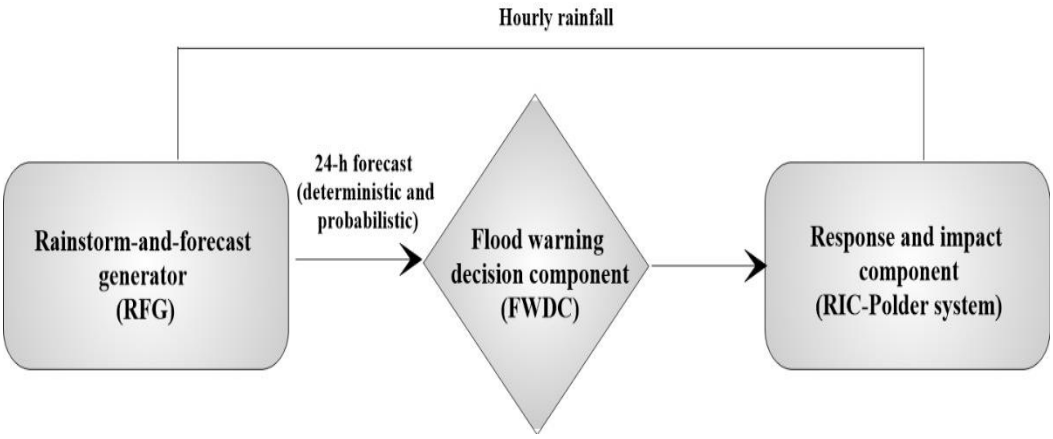


Figure 7.1: Illustration of the framework used in this chapter.

As shall be seen later, and in line with the Monte Carlo (MC) simulation of forecasts in Chapters 5 and 6, the framework's application was based on simulated forecasts generated from a stochastic spatial/temporal rainfall model (RainSim) to demonstrate the benefits that could be derived from the use of forecasts in the management of polder flooding. Therefore, the reader should have in mind that the use of the word "forecast" throughout the chapter refers to a MC simulated forecast which implies that the proposed framework can be applied with real-world forecasts and observed rainfall data.

This chapter is structured as follows: Section 7.2 introduces the research area and the algorithm that has been built to represent the response component of the framework. In section 7.3, the structure of the RFG is explained. In sections 7.4 and 7.5, the rainfall-threshold-based approach adopted to represent the warning decisions in the polder system and decision rules used to simulate that decisions under deterministic and probabilistic forecasts are explained. In sections 7.6 and 7.7, the algorithms developed to represent the pumping strategies and to analyse the operation of the polder system under different forecast scenarios are defined. Finally, in sections 7.8 and 7.9, the results are presented.

7.2 Water Balance Model for the Operation of the Shazou Polder

7.2.1 Research Area

The research area is the Shazhou polder situated in Hexi New Town, located in the southwest of Nanjing City (54.7km^2). This area is surrounded by the Yangtze, Qinhuai, Nanhe, and New Qinhuai rivers (the outer rivers). The topography of Hexi New Town is plain and low lying, lower than the normal water level of the adjacent outer rivers. The town is protected from flooding by embankments and draining stormwater, collected in inner rivers, to the adjacent outer rivers by using pumping stations (Gao *et al.*, 2013)(Figure 7.2).

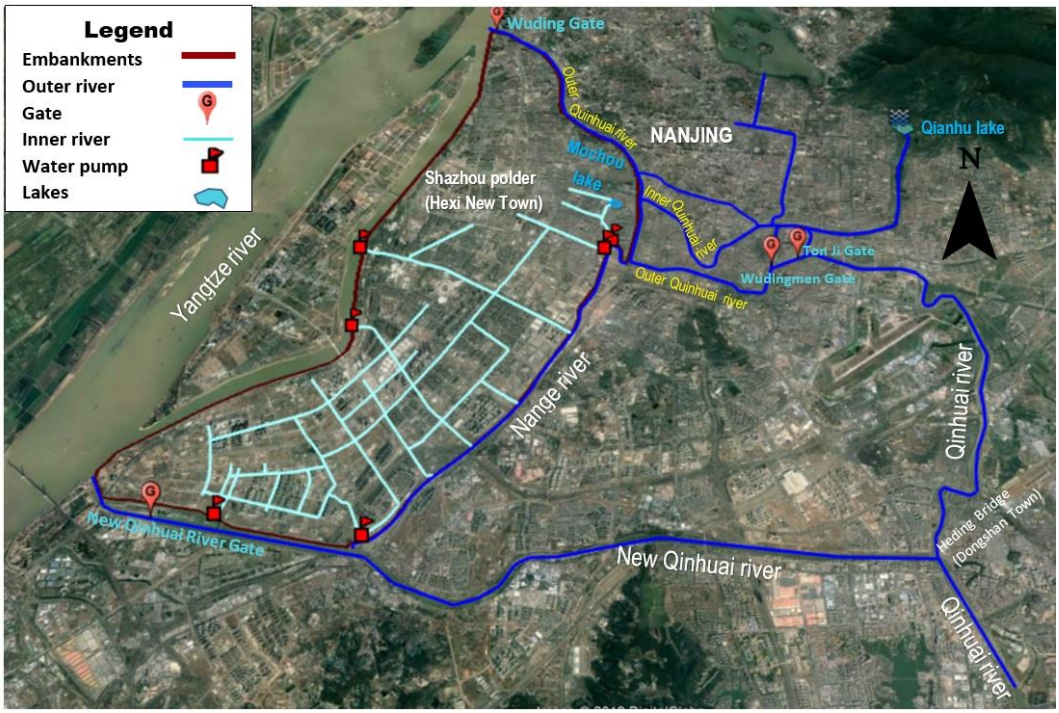


Figure 7.2: Map of the Shazhou polder (Hexi New Town) and its surrounding areas, Nanjing-China

The capacity of the water pumps is described in Appendix A and was provided by Nanjing Hydraulic Research Institute (NHRI).

The Shazhou polder is surrounded on two sides by the Qinhuai river due to its main channel dividing upstream into two branches at Heding Bridge of Dongshan Town. The west branch, the New Qinhuai river, with a total length of 18 km and a previously designed flood capacity of 900 m^3s^{-1}, flows into the Yangtze River via the New Qinhuai river floodgate, whereas the north branch has a length of 22.4 km and a previously designed flood capacity of 600 $m^3.s^{-1}$ (NMG, 2016). The latter is further split into two branches at Tong Ji Gate. One branch passes through Nanjing

city proper and is called the inner Qinhuai river, and the other is known as the outer Qinhuai river. The inner Qinhuai river flows downstream into the outer Qinhuai river, which then flows into the Yangtze River via the Wuding gate. The Wudingmen gate, located just downstream of the Tong Ji Gate, controls the discharges of the inner Qinhuai river, which in turn, receives the discharges of its tributaries coming from the northern area of Nanjing City and the Qianhu lake (Zhao *et al.*, 2017) (Figure 7.2). The aim of the floodgates is to improve the urban water environment and beauty of the city image, which, in the dry season, raises the water level of the Qinhuai River and, in the flood season, releases the urban water flow (Lui and Zhang, 2004).

This research only focuses on the simulation of the water fluxes of the polder system and the inner rivers and neglects any interaction with the adjacent outer rivers.

7.2.2 Water balance model for the operation of the Shazhou polder

In this section, the algorithm used in the framework to represent the water fluxes in the Shazhou polder is explained. As will be seen later, this algorithm will be used for all scenarios considered in this work (the no warning, the perfect forecast, and the imperfect forecast scenarios), which will be modified according to the pumping strategy adopted in each of them.

Owing to the characteristics of polders, the Shazhou polder is assumed in the analysis to be a tank with inputs and outputs (an input-output system). Thus, five processes have been identified to be simulated in the system during a storm: runoff, waterlogging, inflow to inner rivers, pumped water, and storage in the inner rivers.

There are two benchmark water levels during a storm: the water level at the outlet of the pipe hereinafter called the critical water level (h_c), and the initial water level of the inner rivers before the storm arrives (h_o). If the water level exceeds the pipe outlet level, runoff cannot drain from the polder, the *critical condition*. The storage capacity of the inner rivers (S_o^{cap}) is defined by the difference between h_c and h_o (Figure 7.3); expressed in length units (mm). Figure 7.3 shows the system working under non-critical conditions, i.e., when the water level of the inner rivers is below the critical level h_c. However, waterlogging may still occur in this situation if the runoff from the polder exceeds the drainage capacity of the pipe. Currently, a reactive pumping strategy is implemented by the flood managers, in which the rate

of pumping is based on the inflow rate. (The impact of pumping on the outer river levels, which drain to the Yangtze, is not considered.) A critical condition occurs if the pumping is insufficient to prevent the water levels submerging the pipe outlet. Thus, there are two potential states of the system during a storm: *the non-critical condition*, when the water level of the inner rivers (h) is $\leq h_c$, and *the critical condition*, when $h > h_c$.

Under the assumptions detailed above, the system simulation is represented through Eq. 7.1-Eq. 7.7, where the water balance is done at each time step, and all variates are represented in mm.h^{-1}. The model used in this analysis is not a precisely calibrated model, but it does capture the key components of the polder system and has been validated against a historical storm event.

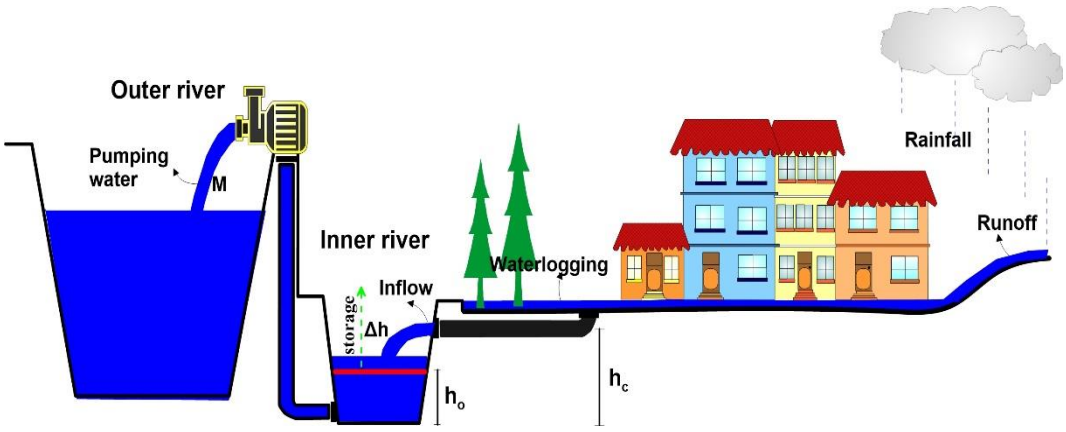


Figure 7.3: Conceptual model of water fluxes in the Shazhou polder system during a rainstorm

This figure shows the scenario when the water level is lower than h_c (non-critical condition) and when q_{max} is surpassed by the inflow. The dashed green arrow indicates the behaviour of the water level of the inner rivers under this condition during the pumping

7.2.2.1 Runoff

The runoff process is represented through Eq. 7.1, which is based on the rainfall-runoff relationship used by Gao *et al.* (2008) in representing the rainfall-runoff process of a neighboring polder.

$$RO_t = 0.55R_t + 0.15R_{t-1} \qquad \text{Eq. 7.1}$$

where R_t [mm] and RO_t [mm] represent the rainfall and runoff value at the time step t [hrs]. As one can note, this equation states that the average runoff coefficient (C_{ru}) in the polder system is 0.7, which can be considered a reasonable value

considering that the impervious area in the polder has been reported to be about 78.5% (Gao *et al.*, 2009). No data were available to recalibrate this relationship for the Shazou polder.

7.2.2.2 Waterlogging

Under non-critical conditions, this process is simulated by using the equation of Gao *et al.* (2008), which uses the capacity of pipe-network drainage as the upper limit to derive the inflow process of inner rivers. Under critical conditions, it is, on the other hand, assumed that the inflow process is blocked and, therefore, the waterlogging cannot be drained. These processes are represented by

$$W_t = \begin{cases} W_{t-1} + RO_t, & if\ (h_t > h_c) \\ max\{0, W_{t-1} + RO_t - r\}, & if\ (h_t \leq h_c) \end{cases} \qquad \text{Eq. 7.2}$$

here r [mm.h^{-1}] is the capacity of the municipal pipe network, h_t [mm] is the water level of the inner rivers at the time step t, W_t is the cumulative excess runoff or waterlogging on the polder at the time step t, and h_c has already been introduced above, which is defined in mm.

7.2.2.3 Inflow

Under non-critical conditions, the inflow to the inner rivers is also represented by using the conceptual model of Gao *et al.* (2008). Under critical conditions, it is assumed that the inflow process is blocked, and the inflow to the inner rivers is null. These processes are represented by:

$$I_t = \begin{cases} 0, & if\ (h_t \geq h_c) \\ min\{r, W_{t-1} + RO_t\}, & if\ (h_t < h_c) \end{cases} \qquad \text{Eq. 7.3}$$

where I_t is the inflow at the time step t.

7.2.2.4 Pumping strategy

The time variable pumping rate (q_t) to be considered in this algorithm will depend on the pumping strategy used to simulate the Shazou Polder. According to the current pump operation of the Shazou polder, the runoff is pumped to the adjacent outer rivers according to the observed inflow *I* (reactive pumping) (Gao *et al.*, 2008, 2009), i.e., $q_t = I_t$, when $I_t < q_{max}$, and when $I_t \geq q_{max}$, the runoff is pumped

at the maximum pumping rate q_{max}. If $h_t > h_c$ (critical condition), it is assumed, based on the analysis of pumping records, that pumping operators will drop the inner rivers to a normal water level (h_n) following the storm event. This normal water level h_n defines the lowest level that pumping operators will draw down the inner rivers, and this only occurs following a critical condition. This strategy and other pumping strategies considered in the framework are explained in detail later on.

7.2.2.5 Storage in the inner rivers

The water storage can be expressed by:

$$S_t = \begin{cases} S_{t-1} - q_{max}, & if\ (h_t \geq h_c) \\ S_{t-1} + max\{0, I_t - q_{max}\}, & if\ (h_t < h_c) \end{cases} \qquad \text{Eq. 7.4}$$

where S_t [mm] is the water storage in the inner rivers at a given time t. Since the variables used in Eq. 7.4 are areal variables (taking the area of the polder as reference), S_t should be understood as a volume of water spreads *over the polder area*. This "areal" value is related to the actual value of the storage as:

$$S_t A_p = S_t^{in} A_{in} \qquad \text{Eq. 7.5}$$

where A_p and A_{in} are the areas of the polder and inner rivers respectively in the same units, and S_t^{in} is the actual value of the water storage in the inner rivers at a given time step t taking mm as unit. If we consider k as the water surface ratio of the polder (ratio between A_{in} and A_p), S_t^{in} can be expressed from Eq. 7.5 as:

$$S_t^{in} = \frac{S_t}{k} \qquad \text{Eq. 7.6}$$

7.2.2.6 The water level in the inner rivers

The water levels of inner rivers in a polder system are usually obtained by simulating the flow processes of the rivers by using, for example, the de St-Venant equations (Liu Jun *et al.*, 2010; Gao *et al.*, 2013). Since this research considers the polder as just an input-output system, these flow processes are not simulated, and the water level in the inner rivers is simply expressed by:

$$h_t = h_{t-1} + S_t^{in} \qquad \text{Eq. 7.7}$$

At the start of each simulation, it is assumed that the initial water level is equal to the normal water level h_n.

7.2.2.7 Inundated area

A relationship is required that expresses the inundated area as a function of water level in the inundated polder. In this sense, an impact curve was assumed for these purposes. This curve is shown in Figure 7.4 and was added to the conceptual model to represent the area inundated in the polder as a function of the depth of water that accumulates in the polder, i.e., the waterlogging. The shape of the function reflects the substantial development in lower areas of the polder, with substantial inundation occurring with initial waterlogging.

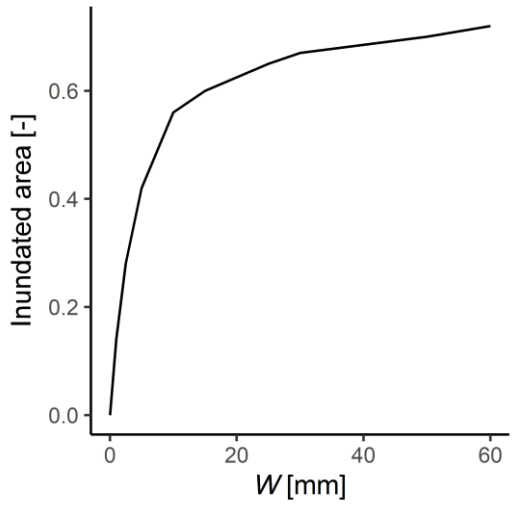


Figure 7.4: Inundated area-waterlogging function
This function was added to the conceptual model to represent the inundated area in the polder system as a function of the depth of waterlogging W [mm]. The inundated area is expressed as a portion of the area of the polder A_p.

7.2.3 Calibration of the model

Particular efforts have been recently made to simulate polder systems in China for flood risk analysis (Gao *et al.*, 2017, 2018; Fang *et al.*, 2018; Wei *et al.*, 2018). However, none of the studies has considered a calibration procedure for the urban drainage system. One of the main reasons could be the lack of data/information. This issue is not the exception in this research; since to calibrate the conceptual model, an average value of the water level of the inner rivers would be needed that

should be derived from several observed points in the river network. In this context, the objective pursued by the calibration process in this research was to represent the conditions that give rise a critical condition situation in the Shazhou polder, rather than the correct simulation of the average observed water level h. For these purposes, this research used the observed rainfall of the 7-July-2016 event that is known to have caused a critical condition in the polder system. This calibration procedure is explained as follows.

Table 7.1 shows the parameter values needed to implement the algorithm for the Shazhou polder (Figure 7.2). Most of the adopted values have been taken from previous studies performed in this case study area (Liu Jun *et al.*, 2010; Gao *et al.*, 2013), while the value of q_{max} was initially obtained by summing the pumping capacity of all pumping stations (Annex 1) and dividing by the polder area (A_p=54.7 km^2). This information has been confirmed based on interaction with the Nanjing Hydraulic Research Institute (NHRI).

Table 7.1 Adopted parameter values for the water balance model of the Shazhou polder.

Parameter	Description	Unit	Value	Source
r	The capacity of pipe-network drainage	mm. h^{-1}	22.14	(Liu Jun et al., 2010)
Δh_n	Difference between the critical and normal water level of the inner rivers, i.e. h_c and h_n, respectively.	mm	500	
k	Water surface ratio	-	0.065	
q_{max}	The maximum pumping capacity for the polder	mm. h^{-1}	9.62 (14.8)	NHRI
C_{ru}	Average runoff coefficient.	-	0.7	Based on Gao *et al. (*2009)

The initial value obtained for q_{max} was taken as a reference, and it was subjected to calibration. The calibration procedure consisted of adjusting the value of q_{max} to produce a critical condition situation for the 7-July-2016 event. The calibrated value of q_{max} (9.62 mm.h^{-1}) represented 65% of the value obtained based on the theoretical maximum(14.8 mm.h^{-1}), which seems reasonable. Figure 7.5 shows the simulated inflow and water level for the 7-July-2016 event. According to the model,

the critical condition started at 5:00 and, and the storage capacity was practically full during the following 4 hours (8:00 am). This result corresponds with observed records, which state that the inundation associated with the 7-July-2016 event occurred around these hours. Thus by using the calibrated value of q_{max} and the additional parameters values shown in Table 7.1, the model simulates the critical condition that was known to occur for that event.

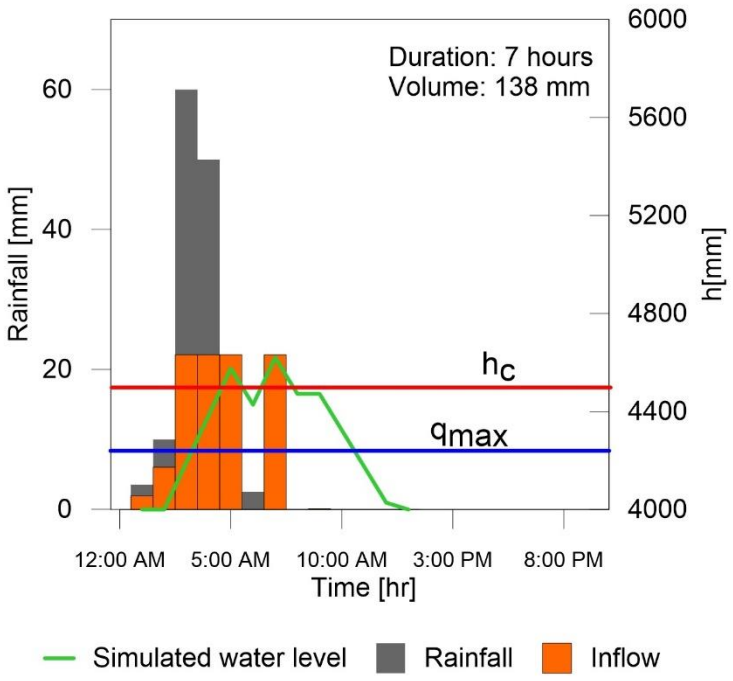


Figure 7.5: Calibration of the conceptual model
This figure shows the simulated water level for the 7-July-2016 rainfall event by using the parameter values shown in Table 7.1. Also shown is the simulated inflow process I_t.

7.3 Monte Carlo generation of rainstorms and their forecasts

The conceptual model explained in section 7.2, which represents the response and impact component RIC of the generic framework (Figure 7.1), has rainfall as the driving input variable. In this section, it is explained how the framework represents the rainstorm-and-forecast generator RFG which uses RainSim V3, a robust and well tested stochastic rainfall field generator (Burton *et al.*, 2008), to represent: i) a daily rainfall (R_{daily}) and its forecast ($\hat{R}_{daily}$), and ii) the hourly rainfall R_t. As will be seen later on, $\hat{R}_{daily}$ will be used in the framework to represent the 24-h forecast in the deterministic forecast scenario, whereas R_t will be used to simulate the water fluxes in the polder system. Note, however, that this framework also needs to represent the 24-h probabilistic forecast, which cannot be derived directly from

RainSim V3. Therefore, to represent this type of forecast, the framework uses a joint distribution of R_{daily} and $\hat{R}_{daily}$ to derive the uncertainty of R_{daily} given $\hat{R}_{daily}$. This section thus first describes the stochastic rainfall field generator and the criteria adopted for calibrating and validating it, and, at the end of the section, the approach used to derive the joint distribution of R_{daily} and $\hat{R}_{daily}$ is explained.

7.3.1 RainSim rainfall field model

The stochastic rainfall modelling in RainSim V3 is based on the Neyman–Scott Rectangular Pulses (NSRP) model, and it can be used for a single site application (a point rainfall generator) or for spatial applications (a spatio-temporal rainfall generator). RainSim V3 operates in three modes: First, the model computes several required statistics from the observed time series; the aim of this stage is to do a statistical characterisation of the rainfall time series. This mode is called *analysis*. Then, the model identifies the parameter set that, according to analytical expectation, best matches the observed statistics. This mode is calling *fitting*. Finally, the model generates synthetic time series using the fitted parameters. This mode is called *simulation*.

7.3.1.1 Spatio-temporal model structure

The spatio-temporal Neyman-Scott Rectangular Pulses (NSRP) model used by RainSim V3 for this application is summarized below (Burton *et al.*, 2008) and illustrated in Figure 7.6. A summary of its parameters is shown in Table 7.2.

- ✓ ***Step 1:*** Storm origins arrive in time in a Poisson process with an occurrence rate λ (Figure 7.6(a));
- ✓ ***Step 2:*** A random number of raincells is generated in space and time for each storm. A parameter β controls the arrival times of the cells after the storm origin; these have an exponential distribution. The centres of the spatially circular raincells are generated by a uniform Poisson process in space with density ρ (Figure 7.7(b)). The radius of each of these raincells is exponentially distributed with parameter γ.
- ✓ ***Step 3:*** Each raincell produces a uniform rainfall rate generating, thus, rectangular pulses. The duration and intensity of each of them are independent and exponentially distributed with parameters η and ξ, respectively;
- ✓ ***Step 4:*** The total rainfall at any time is the sum of all active raincells in time and space.

To account for orography, the rainfall sampled at each site is scaled by a factor ϕ proportional to each site's mean rainfall.

Table 7.2 Input parameters of RainSim V3 for spatio-temporal applications

Symbol	Statistic	Units
λ	1/mean waiting time between adjacent storm origins	(1/h)
β	1/ mean waiting time for raincell origins after storm origin	(1/h)
η	1/mean duration of raincell	(1/h)
ξ	1/mean intensity of a raincell	(h/mm)
γ	1/mean radius of raincells	(1/km)
ρ	Spatial density of raincell centres	(km^{-2})
ϕ	A vector of scale factors, ϕ_m , one for each raingauge	(-)

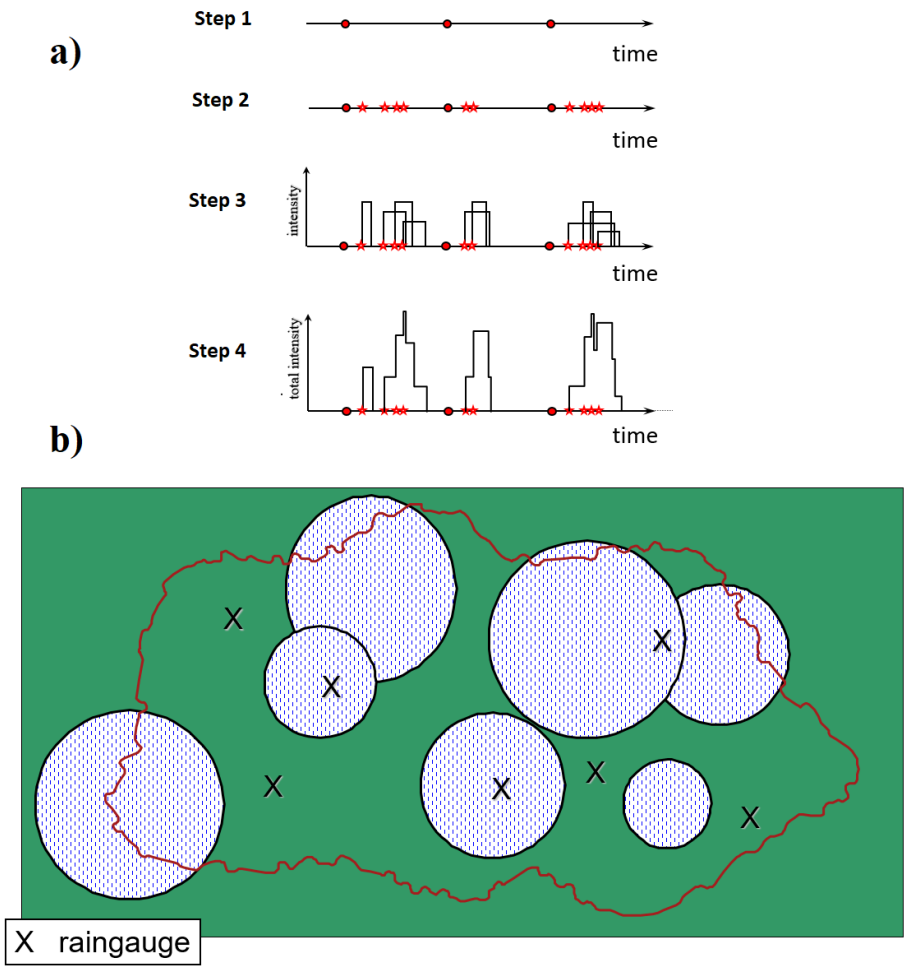


Figure 7.6: Schematic of the Neyman–Scott Rectangular Pulses model used by RainSim V3 in spatial mode

Adapted from Burton *et al.*, (2008)

7.3.1.2 Model fitting procedure

Equations have been developed relating the statistical rainfall model properties to the model parameters (Burton et al. 2008):

$$f_j = f(\lambda, \beta, \rho, \eta, \xi, \gamma, \phi) \qquad \text{Eq. 7.8}$$

where f_j is a model property, j= 1,2, …,m, and m is the number of properties employed in fitting the model. These properties are typically the mean, variance, skewness, autocovariance at a given time lag, probability of an h-hour dry period, wet/wet and dry/dry transition probabilities, and cross-correlations between stations for selected rainfall durations. They can be estimated from the available rainfall time series, and the parameters are then estimated by minimizing the following sum of squares function:

$$SS = \sum_{j=1}^{m} w_j \left(1 - \frac{f_j}{\hat{f}_j}\right)^2 \qquad \text{Eq. 7.9}$$

where $\hat{f}_j$ is an estimated statistic from the observed data, f_j is its model equivalent, and w_j is a preferential weight applied to one or more selected statistics.

7.3.2 Model calibration and validation

7.3.2.1 Available data

RainSim V3 first requires a set of statistics to be computed from an observed sample to provide a statistical characterization of the rainfall time series. These statistics are calculated from hourly and daily data. This subsection describes the sources of the data used for that purpose and other important information.

The daily data were obtained from the Global Historical Climatology Network - Daily (GHCN-Daily) dataset(Menne *et al.*, 2012), which provides a long daily record (62 years) at Nanjing Station (Figure 7.7 and Table 7.3).

The hourly data were collected from the five rainfall stations shown in Figure 7.7. The record length and other important information about the stations are described in Table 7.4.

Table 7.3 Source of the daily records used in RainSim V3.

Station	Length record	Lat.	Long.	Country	Institution	Source
Nanjing	1950-2012	118°43′	32°05′	China	NOAA/NCEI	GHCN-Daily dataset [reference]

Table 7.4: Characteristics of the rainfall stations used in RainSim V3

Code	Name	Lat.	Long.	Record period
62724050	Nanjing	118°43′	32°05′	2012-2016
62935200	Xiaoqiao	118°34′	32°10′	
62936600	Liuhe	118°53′	32°20′	
62936660	Getang	118°44′	32°15′	
63129400	Dongshan	118°51′	31°57′	

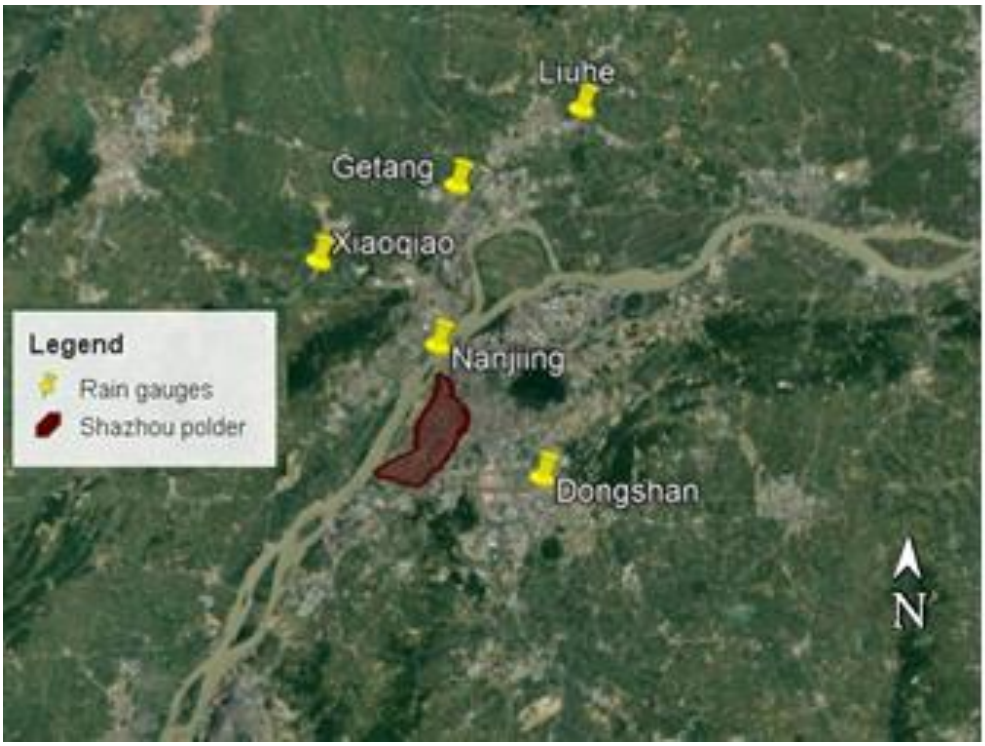


Figure 7.7: Geographical location of the four hourly rainfall stations

7.3.2.2 Results

The calibration (fitting) of the model was done through a numerical optimization based on the procedure described in section 7.3.1.2. The observed statistics used for fitting are shown in Table 7.5. The daily statistics were calculated from the long daily record of the Nanjing station obtained from the GHCN-Daily dataset (Table 7.3), whereas the hourly statistics were calculated from the Dongshan record. The spatial-temporal rainfall model simulated rainfall at five locations (Fig 7.7), with

the same observed statistics at each location, but with different spatial correlation (parameter xcorr in Table 7.5) obtained from the records of the five hourly rainfall stations. (Note, for brevity, the spatial correlation in Table 7.5 is only shown for the Nanjing and Dongshan stations.) This was done based on the criterion adopted to represent the observed rainfalls and their forecasts and will be explained in the next subsection. The fitted parameters are shown in Table 7.6.

Table 7.5 Statistics for RainSim V3.
The statistics correspond to July, which is the rainiest month in Nanjing

Abbreviation	Statistic	Type of information	Observed	Fitted	Weight
mean	The mean *h* hour rainfall accumulation	Daily	6.45	6.42	5
pdyr	The probability that an h hour accumulation is dry, that is strictly less than a specified threshold	Daily	0.69	0.81	6
var	The variance of the *h* hour accumulation	Daily	334.95	334.97	2
corr	The auto-correlation of the *h* hour accumulation of two-time series.	Daily	0.16	0.30	3
xcorr	The cross-correlation of the *h* hour accumulation of two-time series.	Daily	0.90	0.96	2
skew	The skewness coefficient of *h* hour accumulation	Daily	4.86	3.88	3
pdyr	The probability that an h hour accumulation is dry, that is strictly less than a specified threshold	Hourly	0.91	0.93	5
var	The variance of the *h* hour accumulation	Hourly	2.83	2.84	3
skew	The skewness coefficient of *h* hour accumulation	Hourly	11.36	11.28	3

Table 7.6. Fitted parameter for RainSim V3 model

Parameter	Value
λ	0.003967
β	0.077682
ρ	0.001050
η	5.381274
ξ	0.169332
γ	0.015000

The observed and simulated daily annual maximum rainfall are shown by means of a Gumbel plot in Figure 7.8. Each dot represents the maximum daily rainfall recorded in a July in the observed record, and the corresponding vertical lines show the extremes from a 10 member ensemble extracted from the simulation. The range of extremes matches the observed data.

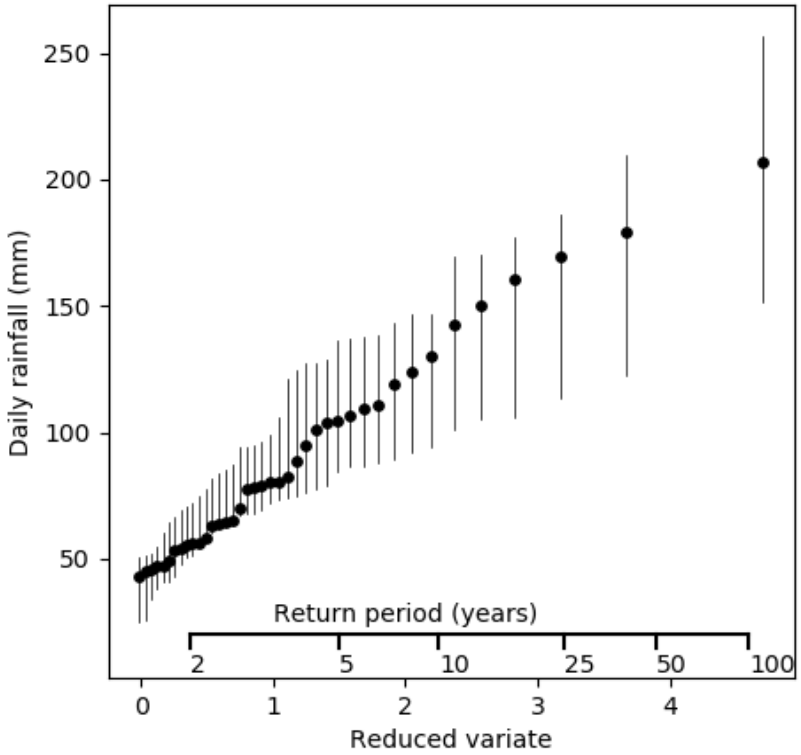


Figure 7.8: A comparison of observed and simulated daily annual maximum for July. The range of the simulated results from a 10 member ensemble is shown at each return period

A scatter plot of the simulated daily rainfall values at Dongshan (63129400) and Nanjing (62724050) is shown in Figure 7.9. As the same statistics are specified at each site during the fitting of RainSim, there is no bias, with the scatter due to the spatial correlation specified between the pair of sites (Table 7.5; xcorr)

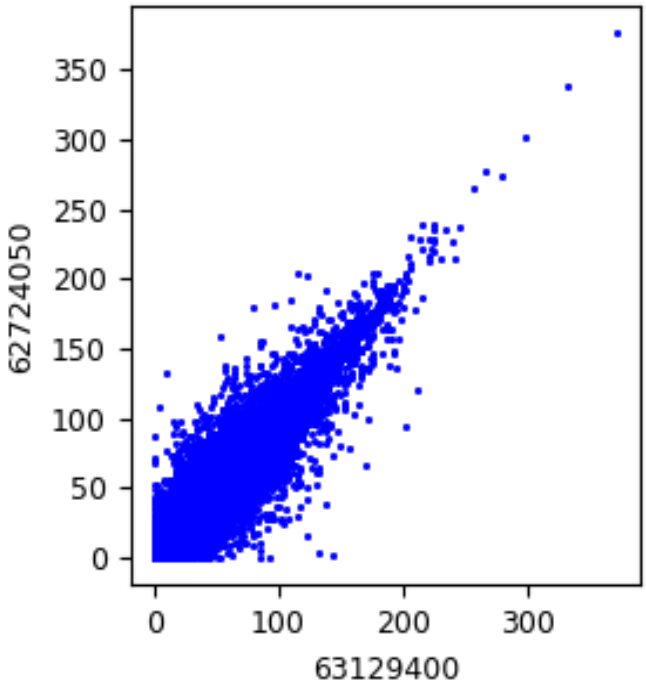


Figure 7.9: Scatter plot of simulated daily rain values (mm) at Dongshan (63129400) and Nanjing (62724050)

7.3.3 Generation of observed and forecast rainfall time series

The rationale for fitting a spatio-temporal model is the following. A relatively simple procedure was needed for generating forecasts of the observed storms generated by a point NSRP model. This was done by selecting one of the five sites to represent observed rainfall and a second site to represent its forecast, but with the same underlying temporal statistics and corresponding parameters, as the observed rainfall site. The cross-correlation between the pair of sites (parameter xcorr in Table 7.5) was used to control the level of agreement between the observed and forecast rainfall time series. Dongshan station was chosen as the location that represents the observed time series and the time series at Nanjing station as its forecast. Since they share the same statistics, there is no bias between them, and the forecast uncertainty is only expressed by the correlation parameter. The Dongshan station was chosen as the observed time series because the records of this station best represented the 7-July-2016 event which was used to calibrate the conceptual

model of the Shazou polder (section 7.2.3). The fitted parameters are shown in Table 7.6. As the means are the same at the two sites, the scale factors in Table 7.2 were the same.

7.3.4 The joint distribution of generated observed 24-h rainfall and their forecasts

The generated 'observed' and 'forecast' time series from the RainSim V3 model can be sampled at any required time interval. As NWP rainfall forecasts are typically available for up to 24 hours in advance, it was decided to adopt a 24 hour (daily) lead time for the forecasts. As explained at the beginning of section 7.3, the 24-h probabilistic forecast to be used in the probabilistic scenario will be derived from the joint distribution of the values of R_{daily} and $\hat{R}_{daily}$ obtained from RainSim V3. This section shows the procedure conducted to i) build this joint distribution and ii) derive the conditional distribution of $\hat{R}_{daily}$ given R_{daily}, i.e., $f(R_{daily}|\hat{R}_{daily})$, which will be used as a measure of predictive uncertainty (PU). To build this joint distribution, a univariate analysis of R_{daily} and $\hat{R}_{daily}$ was first performed. Then the bivariate distribution of R_{daily} and $\hat{R}_{daily}$ was derived. This analysis was carried out generated a time series of 1000 years in length, i.e., with 365000 daily values.

7.3.4.1 Univariate analysis of generated observed daily rainfalls and their forecasts

The univariate analysis of R_{daily} and $\hat{R}_{daily}$ aims to find the best probabilities distributions to represent the marginal distributions in the bivariate analysis. This was done based on the goodness of fit (GoF) of different distributions to values of R_{daily} and $\hat{R}_{daily}$. Since this work focuses on rainfall events that could potentially produce significant runoff events in the polder system, the univariate analysis only considered pair of values that were both greater than 50 mm. After applying this filter, the sample size was reduced to 17,998 daily values. When the sample size is large, it is not suggested basing the GoF on a traditional statistical test (tests based on the p-value)(Tanaka, 1987). Thus, this work used a visual inspection technique to analyse the GoF, which involves plotting the theoretical quantiles against the empirical ones. This plot is known as the quantile-quantile (q-q) plot; If the probability distribution fits the sample data well, the points should fall

approximately along the 1:1 line. The empirical probabilities can be computed by using the Weibull equation (Eq. 4.55).

Among the several probability distributions tested, which were the two-parameter exponential distribution, the three-parameter log-normal distribution, and the three-parameter Gamma distribution, the latter was found to be the best distribution to represent the values of R_{daily} and $\hat{R}_{daily}$. The distribution parameters were computed through the maximum likelihood estimation (MLE) and assuming that the location parameter is known, taken as 50mm for each case. The results for the Gamma distributions are shown in Table 7.7 and Figure 7.10 (the results of the other distributions are not showed to save space in the thesis).

Table 7.7 Estimation of the parameters of the three-parameter gamma distribution for the values of R_{daily} and $\hat{R}_{daily}$

Location parameter		Shape parameter		Scale parameter	
R_o^{daily}	$\hat{R}_o^{daily}$	$\gamma_{R_{daily}}$	$\gamma_{\hat{R}_{daily}}$	$\beta_{R_{daily}}$	$\beta_{\hat{R}_{daily}}$
50	50	1.45	1.44	23.2	23.3

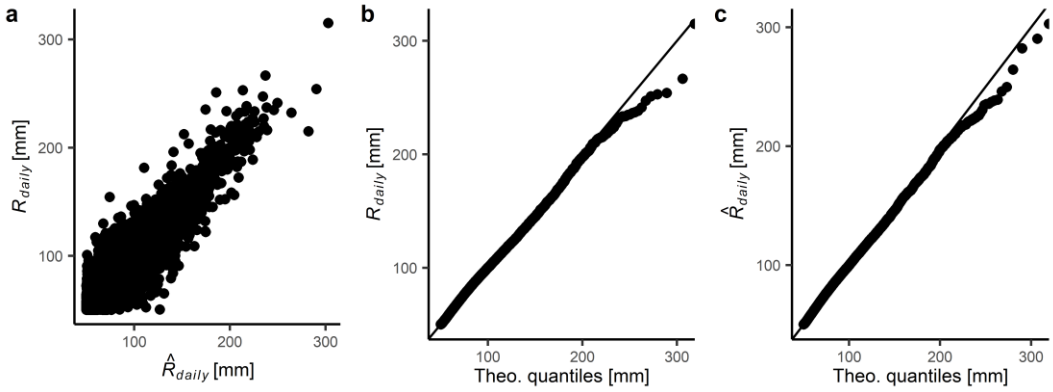


Figure 7.10: Scatter plot of R_{daily} and $\hat{R}_{daily}$ and visual inspection of GoF
Figure a shows the pairs of values of R_{daily} and R_{daily} > 50 mm, and Figure b and c the qq-plots of each of them assuming a three-parameter gamma distribution.

7.3.4.2 Bivariate simulation of observed daily storms and their forecasts

The visual GoF inspection of R_{daily} and $\hat{R}_{daily}$ suggested that the three-parameter Gamma distribution is the best distribution to represent these variables. Thus, one can assume that their bivariate relationship is described by a bivariate Gamma

distribution (BGD). The generator of pairs $(R_{daily}, \hat{R}_{daily})$ and the building of PU expressed as $f\left(R_{daily}\middle|\hat{R}_{daily}\right)$ based on this distribution is explained as follows.

The generator of pairs ($R_{daily}, \hat{R}_{daily}$)

The bivariate generator is based on the Gaussian copula (section 4.4.2). Thus, assuming $(R_{daily}, \hat{R}_{daily}) \sim BGD$, whose marginals are defined by the parameters shown in Table 7.7, pairs $(R_{daily}, \hat{R}_{daily})$ can be simulated based on the following algorithm.

- ✓ **Step 1**: Define the parameters $\beta_{R_{daily}}, \gamma_{R_{daily}}, \beta_{\hat{R}_{daily}}, \gamma_{\hat{R}_{daily}}, R_o^{daily}, \hat{R}_o^{daily}$, and the coefficient of correlation $\rho_{R_{daily},\hat{R}_{daily}}$ in the normal space, i.e., $\rho_{n,\hat{n}}$.
- ✓ **Step 2**: Generate n bivariate normal standardized pairs, i.e., $(\eta, \hat{\eta})$, with correlation $\rho_{n,\hat{n}}$ based on the conditional approach, i.e., building $f(\eta|\hat{\eta}_i)$, and, then, drawing a random value from this conditional distribution.
- ✓ **Step 3**: Compute the CDF of the resulting pairs $(\eta, \hat{\eta})$.
  $u = F(\eta)$ and $r = F(\hat{\eta})$
- ✓ **Step 4**: Compute standardized gamma variables w and $\hat{w}$ associated with the shape parameters $\gamma_{R_{daily}}$ and $\gamma_{\hat{R}_{daily}}$ respectively as:
  $w = F_W^{-1}(u)$ and $\hat{w} = F_{\hat{W}}^{-1}(r)$
  where F_W^{-1} and $F_{\hat{W}}^{-1}$ are the inverse CDFs of the distributions of w and $\hat{w}$ which are in turn defined as:
  $w = \frac{R_{daily} - R_o^{daily}}{\beta_{R_{daily}}}$ *and* $\hat{w} = \frac{\hat{R}_{daily} - \hat{R}_o^{daily}}{\beta_{\hat{R}_{daily}}}$
  where, as was explained above, R_o^{daily} and $\hat{R}_o^{daily}$ and $\beta_{R_{daily}}$ and $\beta_{\hat{R}_{daily}}$ are the location and scale parameters of R_{daily} and $\hat{R}_{daily}$, respectively.
- ✓ **Step 5**: Pairs $(R_{daily}, \hat{R}_{daily})$ can be finally obtained as:
  $R_{daily} = \hat{R}_o^{daily} + \beta_{R_{daily}} w$
  $\hat{R}_{daily} = \hat{R}_o^{daily} + \beta_{\hat{R}_{daily}} \hat{w}$

To compute the value of $\rho_{R_{daily},\hat{R}_{daily}}$ in the normal space, i.e., $\rho_{n\hat{n}}$, one can first find the relationship between the correlation coefficient in the normal (ρ_N) and gamma (ρ_{Gam}) space associated with the parameter set $\boldsymbol{\theta}_{R_{daily},\hat{R}_{daily}}$. It is performed by following the steps of the above-mentioned algorithm, assuming several arbitrary values of ρ_N in Step 2, and computing the associated value of ρ_{Gam} numerically. For any specified value of ρ_{Gam}, the value of $\rho_{n\hat{n}}$ can then be obtained via interpolation based on the pairs of values (ρ_{Gam}, ρ_N).

The building of PU

The framework uses the above-mentioned bivariate gamma generator to build the conditional distribution of $f\left(R_{daily}\middle|\hat{R}_{daily}\right)$ which will be used to provide a measure of PU in the probabilistic forecast scenario. This is done by simply sampling n values from the conditional distribution $f(\eta|\hat{\eta}_i)$ obtained in the bivariate-gamma generator (Step 2) and then conveying the resulting values to gamma space by following the steps of the algorithm in terms of R_{daily}. PU can be finally expressed in terms of density values by computing the kernel density estimations of the sampled values.

Example of application

The value of $\rho_{R_{daily},\hat{R}_{daily}}$, i.e., the correlation in the Gamma space ρ_{Gam}, was estimated through the sample correlation coefficient of the pair of values shown in Figure 7.10a. Table 7.8 shows the seven parameters of the BGD of R_{daily} and $\hat{R}_{daily}$.

Table 7.8 Values of the parameters of bivariate gamma distribution of R_{daily} and $\hat{R}_{daily}$

R_o^{daily}	$\hat{R}_o^{daily}$	$\gamma_{R_{daily}}$	$\gamma_{\hat{R}_{daily}}$	$\beta_{R_{daily}}$	$\beta_{\hat{R}_{daily}}$	$\rho_{R_{daily},\hat{R}_{daily}}$
50	50	1.45	1.44	23.2	23.3	0.93

The results of the bivariate simulation based on the generator explained above are illustrated in Figure 7.11. Figure 7.11a shows the joint density of 50,000 pairs $(R_{daily}, \hat{R}_{daily})$ from the the bivariate gamma distribution. Figure 7.11b shows an example of the PU expressed as $f\left(R\middle|\hat{R}_{daily}\right)$ for a forecast value $\hat{R}_{daily_i}$ shown in Figure 7.11a (grey dot). Note that in the framework $\hat{R}_{daily_i}$ is obtained from RainSim and its associated $f\left(R\middle|\hat{R}_{daily}\right)$ will be derived from the BGD.

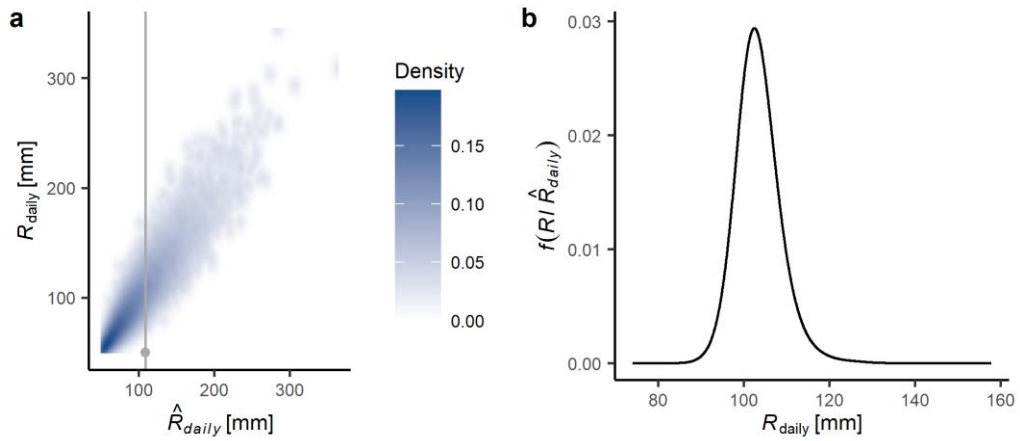


Figure 7.11: Results of the bivariate simulation of R_{daily} and $\hat{R}_{daily}$

In this figure, a shows the joint density of 50,000 pair of values of the bivariate gamma distribution defined by the values of the parameters shown in Table 7.8; the pairs of values were obtained through the bivariate gamma generator. Figure b shows the conditional distribution $f(R_{daily}|\hat{R}_{daily})$ associated with the forecast value $\hat{R}_{daily_i}$ represented by a grey dot in Figure a. $f(R_{daily}|\hat{R}_{daily})$ is obtained by slicing the joint distribution through the grey line that crosses the forecast value (grey dot).

7.4 Daily rainfall forecast thresholds for the polder system

In the previous section, the procedures for deriving the deterministic and probabilistic 24-h rainfall forecasts within the RFG component of the framework were described. As will be seen in the next section, storm warning decisions based on these forecasts will be represented through a rainfall-threshold approach. Thus, in this section, the methodology adopted to derive the daily rainfall thresholds in the Shazou polder is described. As will be seen below, this methodology considers the uncertainty of the profile of the daily rainfall and the initial condition of the water level in the inner rivers at the time the forecast is issued h_o.There are few works that have considered the uncertainty of rainfall characteristics when analysing rainfall thresholds for flood warning systems (Wu *et al.*, 2015), and, to the best of my knowledge, a methodology to quantify rainfall thresholds for flood warning purposes in a polder system has not been published to date. The methodology used in this work is explained as follows.

A daily rainfall threshold (RT_{daily}) for a polder system can be defined as the volume of a daily rainfall which brings the water level of the inner rivers to the critical level h_c, i.e., a daily rainfall that fills up the storage capacity of the inner rivers. Thus, daily rainfall values greater than RT_{daily} falling on the polder area bring the inner rivers to critical conditions. In this context, RT_{daily} has to be associated with h_o,

i.e., the initial condition of the water level of the inner rivers at the time the forecast is issued. Note also that a critical condition situation depends not only on the volume but also on the rainfall profile of the storm, causing critical conditions in the polder. Daily rainfalls with significant rainfall volumes spread uniformly during the day might not cause critical conditions in the polder due to the runoff rate reaching the inner river being equal to or lower than the pumping capacity of the polder q_{max}. In this case, water is not stored in the inner rivers, and the polder manager can drain the runoff smoothly if he pumps in proportion to the drainage from the polder. However, other daily rainfalls with similar rainfall volumes, but concentrated in relatively short time periods, might cause critical conditions in the polder system due to the runoff rate might be higher than the maximum pumping capacity q_{max}. In this case, the water is pumped at a rate equal to q_{max}, but the water level rises, and a critical condition situation can be reached. To account for this uncertainty, the stochastic rainfall model explained in section 7.3 provides daily rainfall profiles based on the generated hourly values of those events that could potentially produce significant runoff events in the polder system (the 17,998 profiles associated with the R_{daily} values are shown in Figure 7.10a). A value of RT_{daily} can be computed for each profile using a trial-and-error approach with the water balance model of the Shazhou polder described in section 7.2 with the pumping strategy that represents the current reactive pump operation in the Shazou polder (section 7.2.2.4). Then, RT_{daily} is estimated as a p-quantile of the resulting PDF of these values. The algorithm can be summarized as follows:

- ✓ ***Step 1:*** From the RainSim V3 simulations, define observed daily rainfalls that could potentially produce significant runoff events in the polder system (daily rainfalls > 50 mm).
- ✓ ***Step 2:*** Define different initial conditions as:

$$h_o^j = h_n + j\frac{\Delta h_n}{n_o} \qquad \text{Eq. 7.10}$$

where h_o^j is the initial condition j; n_o is the number of initial conditions considered, h_n is the normal level, and Δh_n is the difference between the critical water level h_c and the normal water level h_n, here taken as 0.5m with h_n= 4000 mm.

- ✓ **Step 3:** For each h_o^j, perform the following sub-steps:
  a) By using the conceptual model of the polder system described in section 7.2 and the pumping strategy that describes the current

pump operation in the Shazou polder (section 7.2.2.4), obtain values of RT_{daily} by rescaling all the values of the daily rainfalls obtained in Step 1 to make them larger or smaller until the resulting water level of the inner rivers hits the critical level h_c.

b) Define the PDF of RT_{daily}, i.e., $f(RT_{daily})$, with the values obtained in sub-step a.

c) Define the rainfall threshold associated with h_o^j, i.e., RT_{daily}^j, as the p-probability quantiles of $f(RT_{daily})$.

Thus, if we have an initial condition h_o^j in the inner rivers at the time the 24-h forecast of rainfall is issued, and that forecast is greater than RT_{daily}^j, a critical condition will be reached, and a proactive response action should be conducted to avoid the inner rivers reaching the critical level.

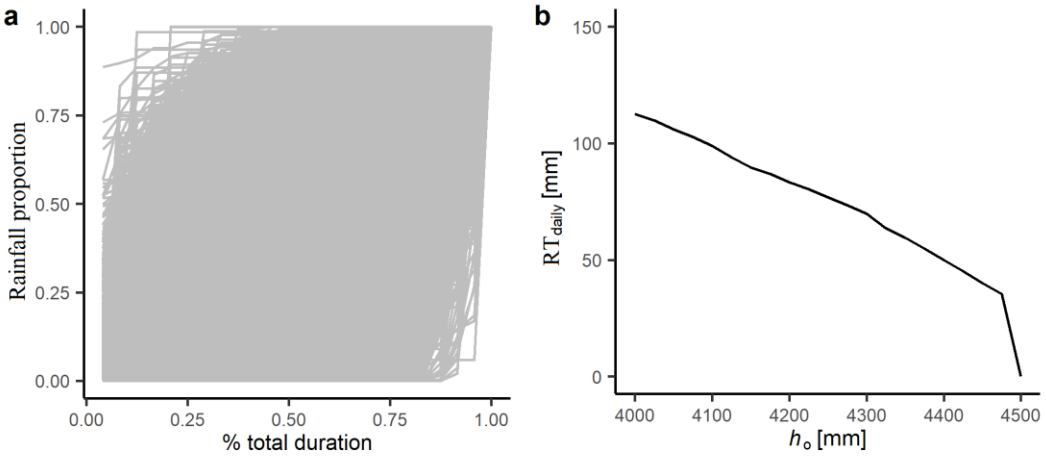


Figure 7.12: Rainfall thresholds for the Shazhou polder

Figure a shows the dimensionless mass curve of the proportion of observed daily rainfalls considered to obtain RT_{daily} values. Figure b shows the values of h_o^j against RT_{daily}^j. The value of RT_{daily} are computed as the 0.01-probability quantile of $f(RT_{daily})$

The results of applying the above-mentioned approach are illustrated in Figure 7.12, which considers twenty initial conditions ($n_o = 20$ in Eq. 7.10). Figure 7.12a shows the 17,998 daily rainfall profiles obtained from RainSim V3 as dimensionless mass curves. Figure 7.12b shows the values adopted for $RT_{daily}^j$by assuming them as the 0.01-probability quantile of $f(RT_{daily})$. By doing that, one expected to remove all the uncertainty with respect to the rainfall profile. Thus, the framework assumes that the values greater than these quantiles will bring the inner rivers to a critical condition. Finally, it is worth noting that the rainfall thresholds for the normal condition h_n= 4000 mm is > 100 mm which corresponds to the warning categorised as "yellow" for Nanjing. A yellow warning is the second

lowest on the 4 colour coded rainstorm warning system used in Nanjing; it triggers a consultation meeting headed by the Commander of Flood Control.

7.5 Warning decisions for the polder system based on 24-h forecasts

A storm warning aims to provide time in advance to the polder manager to conduct a proactive action in the polder system to avoid a critical condition situation. As was mentioned above, the framework considers a storm warning based on a 24-h total rainfall forecast. Figure 7.13 shows the chronology adopted by the framework for the operation of the polder system by considering this type of warning system. It is assumed that the warning is issued at midnight, and, thus, the polder manager can conduct a proactive strategy based on the 24-h forecast. Note that the end of the proactive action depends on the pumping strategy adopted by the polder manager and where t_{pro} is located, and the storm might arrive before or after the proactive period (the types of pumping strategies considered in this work are explained in the next section). The warning decision made at midnight can be based on deterministic and probabilistic rainfall thresholds. The framework represents those decisions based on the values of daily rainfall thresholds RT_{daily} obtained in the prior section. This is explained as follows.

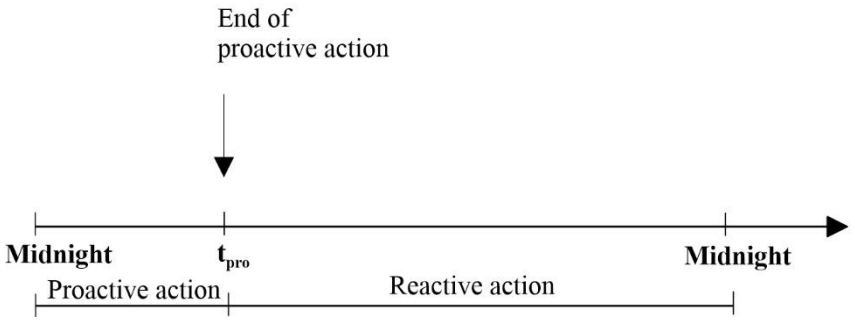


Figure 7.13: Chronology of the operation of the polder system by considering a storm warning based on a 24-h forecast horizon
This framework assumes that storm warnings are issued at midnight; after that, the polder manager can conduct a proactive pumping action. The end of this action, designated t_{pro} in this figure, depends on the strategy adopted by the polder manager. The storm might arrive before or after t_{pro}.

7.5.1 Deterministic warning decisions

This warning decision is based on the deterministic-24h forecasts $\hat{R}_{daily}$, generated from RFG, and the daily rainfall threshold values, (RT_{daily}). A warning is issued if $\hat{R}_{daily}$ is greater than RT_{daily}. The decision rule of this warning decision is given by Eq. 7.11 and illustrated in Figure 7.14.

$$Det.warning\ decision = \begin{cases} if\ (\hat{R}_{daily} > RT_{daily}) & warn, \\ else, & not\ warn \end{cases} \quad \text{Eq. 7.11}$$

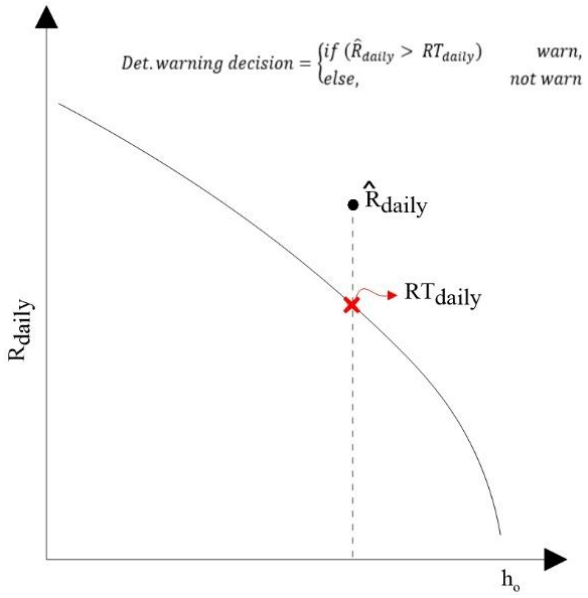


Figure 7.14: Illustration of the deterministic warning decision based on a daily rainfall threshold

This figure shows the illustration of the decision rule used to represent the deterministic warning decision described by Eq. 7.11. Note that RT_{daily} is associated with h_o

7.5.2 Probabilistic warning decisions

The probabilistic warning decision considers the probabilistic-24h forecast and a probabilistic threshold (PT). The former has to do with the exceedance probability (PE) of the daily rainfall threshold RT_{daily}, i.e., the probability of having an observed daily rainfall volume R_{daily} greater or equal than RT_{daily}. PE is obtained through the conditional distribution of R_{daily} given $\hat{R}_{daily}$, i.e., $f\left(R_{daily}|\hat{R}_{daily}\right)$, derived from the joint probability of R_{daily} and $\hat{R}_{daily}$ (section 7.3.4). PT is a probabilistic threshold value on which the warning decision is based. A warning is issued if PE is greater than PT, the latter is a value to be analysed in the framework. The decision rule of this warning decision is shown in Eq. 7.12 and illustrated in Figure 7.15.

$$Probabilistic\ decision = \begin{cases} if\ (PE > PT) & warn, \\ else, & not\ warn \end{cases} \quad \text{Eq. 7.12}$$

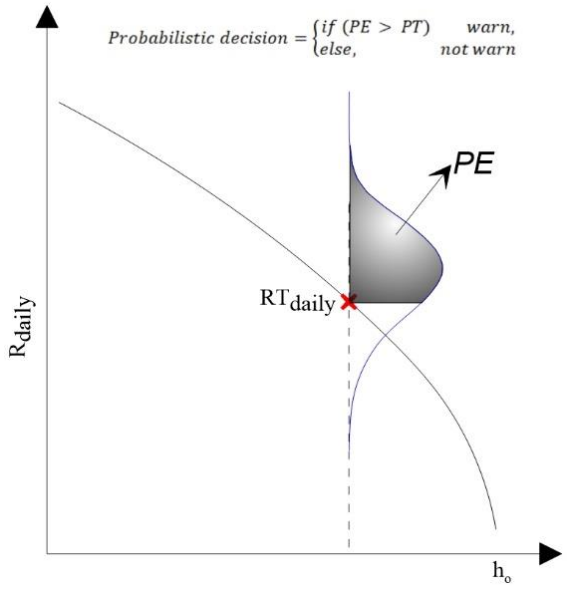


Figure 7.15: Illustration of the probabilistic warning decision based on a daily rainfall threshold

This figure shows the illustration of the decision rule used to represent the probabilistic warning decision described by Eq. 7.12. The probabilistic threshold *PT* is a value to be analysed in the framework

7.6 Pumping strategies under different forecast scenarios.

In the prior section, we defined the probabilistic and deterministic warning decisions that should drive a proactive pumping strategy in the polder system. In this section, the representation of these pumping strategies in the framework will be explained. Here, we also explain the pumping strategies adopted for the no warning and perfect forecast scenarios, which will be considered as the two benchmark cases in the analysis of the framework. Before explaining these pumping strategies, this section starts by explaining two important concepts: i) the water balance of an observed daily runoff causing a critical condition in the polder, and ii) the proactive criterion used to represent a proactive pumping strategy. Understanding these two concepts is important to comprehend the rationality adopted in the pumping strategies.

7.6.1 Water balance of an observed daily runoff causing critical conditions under a reactive action

If one wishes to analyse/add a proactive action in the current pumping operation of the Shazhou polder (where pumping actions are mainly driven by the observed

inflow to the inner rivers) to avoid a critical condition situation, one should first understand the water balance of the polder for an observed daily runoff causing that condition under that pumping strategy (reactive strategy). This helps to identify target variables that should be considered in the proactive strategies to avoid a critical condition in the inner rivers. This water balance will be explained in terms of areal values and length units. The values of volumes can therefore be understood as the depth of the volume of water spread over a reference area, e.g., the area of the polder or the inner rivers. Hence, the water balance of a critical observed daily runoff can be expressed by:

$$RO_{daily}^{c} = V_p^c + S_o^{cap} + S_c \qquad \text{Eq. 7.13}$$

where RO_{daily}^{c} is the observed daily runoff causing critical conditions, V_p^c is the portion of this critical runoff reactively pumped during the storm to the adjacent outer rivers, S_o^{cap}, as was explained in section 7.2.2, is the storage capacity of the inner river before the storm arrives, and S_c is the portion of the critical runoff that brings the water level of the inner rivers beyond the critical level. Figure 7.16 shows a conceptualization of the water balance explained above. This figure illustrates how the volume of a critical runoff RO_{daily}^{c} is split into the drainage system according to the current pump operation of the Shazou polder (reactive one, see section 7.2.2.4). When the critical storm starts, the initial condition of the water level of the inner rivers is h_o and the storage capacity of the inner rivers is S_o^{cap}. V_p^c is the volume of the critical runoff drained to the adjacent outer rivers during the critical storm, and S_c is the volume of this critical runoff that brings the water level of the inner rivers beyond the critical level. As one can note, S_o^{cap} is the only variable known before the storm arrives, and V_p^c and S_c are, therefore, the target variables in a proactive pumping strategy. To know the values of these variables, one should have perfect knowledge of the profile and volume of the coming daily rainfall causing critical conditions in the polder. If the polder manager would have perfect knowledge about these two variables, he or she could pump S_c before the storm and V_p^c during the storm; the resulting scenario would then be that the storage capacity of the inner rivers would be exactly full at the end of the storm, thus avoiding a critical condition situation.

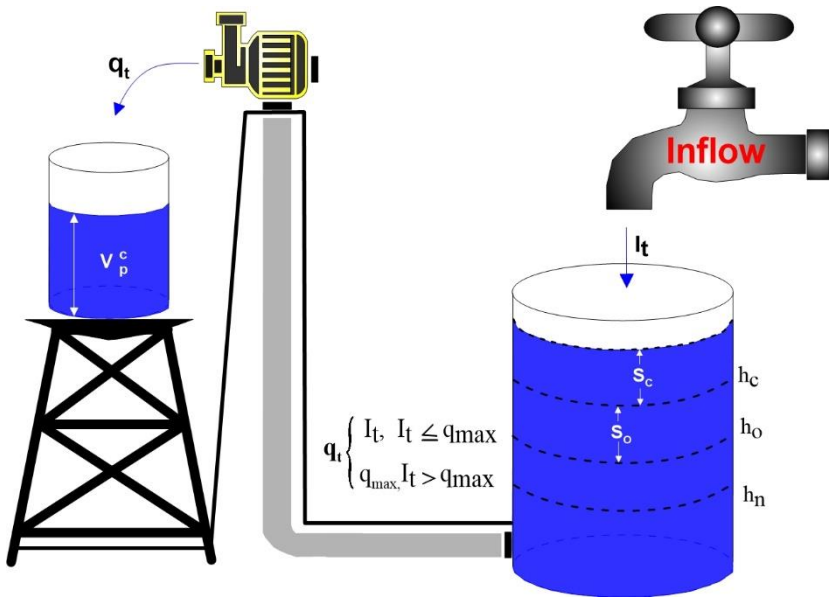


Figure 7.16: Conceptualization of the water balance of an observed daily runoff causing critical conditions in the polder system based on a reactive pumping strategy

According to the current pump operation of the Shazou polder, the runoff is mainly pumped to the adjacent outer rivers according to the observed inflow I (reactive pumping), i.e., $q_t = I_t$, when $I_t < q_{max}$, and when $I_t \geq q_{max}$, the runoff is pumped at a pumping rate q_{max} (section 7.2.2.4). The water level of the inner rivers starts to rise in the latter condition.

7.6.2 Criteria for the proactive pumping strategy

In the prior section, the target variables to be considered in a proactive pumping strategy were identified, i.e., V_p^c and S_c. Now, one should adopt a proactive criterion to pump the volume of water associated with those variables to avoid a critical condition situation. For example, assuming perfect knowledge of the values of these variables, one criterion could be pump V_p^c and S_c together once the storm arrives. Note that, even for the perfect-knowledge assumption, this strategy could not be a good one since the inflow I of some storms can exceed the pumping capacity of the polder q_{max} during all the storm (this is the case for high-intensity storms), leaving no capacity to pump S_c during the storm. Thus, another valid criterion could also be to try to pump S_c, or a portion of it, before the storm arrives, and then wait for the storm to arrive before completing the pumping strategy. This proactive criterion has been adopted in this framework to represent the proactive pumping strategy under the perfect and imperfect-forecast scenarios, where the former is based on the true values of these variables, and the latter use their forecasts.

Thus, the proactive pumping strategies explained later will be made up of a proactive and reactive action. The proactive action has to do with the volume of water pumped before the storm arrives (V_p^{before}), and the reactive action has to do

with the volume of water pumped during the storm (V_p). The proactive pumping strategy, *P.strategy*, can be, therefore, defined by:

$$P.strategy = \begin{cases} V_p^{before} & t \leq t_{pro}; \\ V_p & t \geq \max(t_{arrive}, t_{pro}); \end{cases} \quad \begin{matrix} proactive\ action \\ reactive\ action \end{matrix} \qquad \text{Eq. 7.14}$$

where t_{arrive} is the time at which the observed storm arrives in the polder and t_{pro} is the proactive action period (Figure 7.13). Thus, several pumping rules can be set based on this proactive criterion and under different types of forecast information. This proactive criterion will be the basis for the proactive pumping strategies considered in the framework.

7.6.3 Reactive strategy – no forecast

Based on the case study, a proactive pumping strategy can be understood as a proactive pumping action that can be added to the current operational pumping procedure, which can be considered as a reactive one. Thus, before introducing the proactive pumping strategies, an explanation is provided of how the current reactive pumping actions conducted in the Shazou polder for the critical and no critical conditions will be simulated in the framework.

A reactive pumping strategy can be defined as a pumping action driven by the inflow of the inner rivers I. Under non-critical conditions, this pumping strategy is represented by the following operational principle (Gao *et al.*, 2008, 2009): When the water level starts to rise,

If: the inflow exceeds the pumping capacity of the polder system q_{max},

pump the water at the latter rate, while the excess water is stored in the inner rivers, raising the water level

Else: pump the water at the inflow rate I_t.

For the critical condition, it is assumed that the maximum pumping capacity q_{max} is used in the polder. It is also considered that, after the critical condition has been reached and the inflow has stopped, the polder manager drops the water level of the inner river to the normal level h_n by using the maximum pumping capacity q_{max}. Furthermore, if the resulting water level of the inner rivers after the storm is below

the critical level h_c, it is assumed that the polder manager keeps the water level of the inner rivers at that level. These principles are represented by:

$$q_t = \begin{Bmatrix} q_{max}, & if\ (h_t > h_c)\ until\ h = h_n \\ min\{q_{max}, I_t\}, & if\ (h_t \leq h_c\ \&\ I_t > 0) \\ 0, & if\ (h_t < h_c\ \&\ I_t = 0) \end{Bmatrix} \quad Eq.\ 7.15$$

Where, as was explained in section 7.2.2, q_t [mm.h^{-1}] is the pumping rate at time step t; the other variables have also been introduced in this section.

The adopted reactive pumping strategy has three mains assumptions: i) the pumping starts when a storm starts –reactive action-, ii) once the dropped water level reaches a given water level – here assumed as h_n-, the pumping ends, and iii) the water level of the inner rivers can be higher than the level that defines the end of the pumping. This behaviour has been observed in the operation of the Shazhou polder (Song, 2019).

7.6.4 Proactive strategy under perfect forecast information

In this section, one of the two proactive pumping strategies analysed in the framework is explained. This proactive pumping strategy is designed under the assumption of a perfect forecast and is the best scenario when simulating the polder system as one is assuming perfect knowledge about the target variables S_c and V_p^c.

Adopting the proactive criterion described in Eq. 7.14, V_p^{before} in this equation should be equal to S_c; and then, when the storm arrives, V_p will be equal to V_p^c. Note, however, that, when applying this strategy, the storage capacity of the inner rivers at the end of the 24hr period will be full, which would produce a critical condition situation for the next day, even for a weak storm. To avoid this, V_p^{before} under this scenario is expressed by:

$$V_p^{before} = S_c + (h_c - h_{ref}) \quad Eq.\ 7.16$$

where S_c is expressed as length units (section 7.6.1), and h_{ref} is a reference level of the inner rivers. The reference level h_{ref} is the level at which one wants the water level to be at after the pumping actions; it must be neither too high nor too low. In this first case, a critical condition situation can be produced in the next day,

even by a weak storm. In the second case, the strategy can be considered expensive. When considering the value of V_p^{before} based on this equation, one makes sure that the level of the inner rivers after the end of the storm will be equal to h_{ref}.

In this strategy, the volume of water pumped during the reactive action, i.e., V_p^c, is simulated by *Eq. 7.15*, i.e., the current pump operation of the polder system. For the proactive action, it is assumed that the polder manager pumps a volume of water equal to S_c+$(h_c - h_{ref})$ with a pumping rate equal to q_{max}.

S_c should be computed prior to the analysis of the daily storm by performing in advance the 24h-water balance of the polder system based on the current pump operation of the polder system, i.e., reactive pumping actions, and using the observed profile and volume of the daily rainfall to be analysed (perfect forecast). It can be computed through the following algorithm.

- ✓ ***Step 1***: Assume the polder system to be a tank - an input-output system - and compute the hourly runoff RO_t by using Eq. 7.1, and its associated waterlogging W_t and inflow I_t through Eq. 7.2 and Eq. 7.3, respectively, for the no critical condition situation.
- ✓ ***Step 2***: Compute the hourly water storage as:

$$S_t = \begin{cases} 0, & if\ (I_t \leq q_{max}) \\ S_{t-1} - (q_{max} - I_t), & if\ (I_t > q_{max}) \end{cases} \qquad \text{Eq. 7.17}$$

- ✓ ***Step 3***: Compute the maximum value of S_t, i.e., $S_{max}(S_1, S_2, S_3, \dots S_{24})$, and compute S_c as:

$$S_c = S_{max} - S_o^{Cap} \qquad \text{Eq. 7.18}$$

The chronology of the perfect forecast pumping strategy can be summarized as follows.

- ✓ At midnight, the value of S_c is delivered to the polder manager, and the polder manager conducts the proactive action by pumping a volume of water equal to $S_c + (h_c - h_{ref})$ (Eq. 7.16) with a pumping rate = q_{max}.
- ✓ Then, the polder manager waits for the arrival of the storm. If the storm arrives before $S_c + (h_c - h_{ref})$ has been pumped, the manager will continue with the proactive strategy, and also implement the reactive strategy. Here, the pumping rate used is q_{max} until the target volume has been pumped.

- ✓ Finally, the polder manager completes the pumping strategy by conducting the reactive action once the storm arrives, which is represented by Eq. 7.15. The volume of water pumped during the reactive period will be equal to V_p^c and the level of the inner river at the end of the storm will be equal to h_{ref}.

Note that the representation of the perfect forecast pumping strategy does not mean that the polder system will not be affected by waterlogging. There are two conditions causing waterlogging under the perfect forecast scenario:

- ✓ ***Condition 1***: When the runoff rate RO overcomes the capacity of the drainage system r (Eq. 7.2).
- ✓ ***Condition 2***: When the runoff starts at midnight, and the inflow overcomes the pumping capacity of the polder system q_{max}, i.e., before the proactive strategy can be implemented. Under this condition, t_{arrive} in Eq. 7.14 is zero, and the proactive action cannot be conducted. In this case, the polder manager does not have response capacity for the critical storm, and he/she is only able to use a pumping rate equal to q_{max}, whereas the water level of inner rives rises until a critical condition situation is reached.

7.6.5 Proactive strategy under imperfect forecast information

In the prior section, it was explained how the framework will represent the best scenario when simulating the polder system, i.e., the perfect forecast scenario, so there is no uncertainty about the target variables S_c and V_p^c. Here, a proactive pumping strategy based on imperfect forecast information is explained, where the uncertainty of the target variables is fully considered. This proactive pumping strategy can be applied by considering probabilistic and deterministic forecast information and will be the response to the deterministic and probabilistic warning decisions explained in section 7.5. It will, therefore, be used for both scenarios in the analysis of the framework. This proactive pumping strategy is detailed below

There are a number of proactive strategies one can design to operate the polder system under the proactive criterion described in Eq. 7.14. In this work, only one of them has been used to describe a proactive strategy under different levels of imperfect forecast information. This strategy was called Type-1 pumping strategy. The reactive pumping action is represented by *Eq. 7.15* (the current pump operation of the polder system), and the proactive pumping action, which can be based on deterministic and probabilistic forecasts, is a pumping action driven by a storm warning, i.e., it is only conducted if a storm warning is issued.

The Type-1 pumping strategy assumes that, once a storm warning is issued, the polder manager has a forecast of the total critical daily runoff RO_c^{daily} (Eq. 7.13), designated by $\widehat{RO}_c^{daily}$, but has no knowledge of the daily rainfall profile. Thus, the polder manager has to deal with the proactive action based on knowledge of $\widehat{RO}_c^{daily}$ and S_o^{Cap}; the latter is assumed known because he/she knows the storage capacity of the inner rivers before the storm arrives. This equation can be expressed in terms of the true values of these two variables as:

$$RO_c^{daily} = V_{excess} + S_o^{Cap} \qquad \text{Eq. 7.19}$$

where V_{excess} is the portion of the critical daily runoff expressed by $V_p^c + S_c$ in Eq. 7.13. Based on Eq. 7.19 and the information assumed known by the polder manager; one can, therefore, say that he/she has an estimate of V_{excess} which can be computed as:

$$\hat{V}_{excess} = \widehat{RO}_c^{daily} - S_o^{Cap} \qquad \text{Eq. 7.20}$$

where $\hat{V}_{excess}$ is the estimate of V_{excess}. Thus, the estimate of S_c in Eq. 7.13 can be computed as a portion of $\hat{V}_{excess}$ as:

$$\hat{S}_c = \alpha \hat{V}_{excess} \qquad \text{Eq. 7.21}$$

where $\hat{S}_c$ is the estimate of S_c and α is a parameter with a value between 0 and 1 and represents the portion of $\hat{V}_{excess}$ that represents $\hat{S}_c$, i.e., α represents the proactive pumping factor in the pumping strategy. Based on Eq. 7.21, the estimate of V_p^c is given by:

$$\hat{V}_p^c = (1 - \alpha)\hat{V}_{excess} \qquad \text{Eq. 7.22}$$

Thus, the Type-1 pumping strategy assumes that V_p^{before} in Eq. 7.14 should be equal to $\hat{S}_c$ (Eq. 7.21); and then, when the storm arrives, it assumes that V_p will be equal to $\hat{V}_p^c$ (Eq. 7.22).

To summarize, $\hat{S}_c$ represents the estimate of the portion of a critical observed daily runoff that brings the water level of the inner rivers beyond the critical level h_c. Since the forecast of the rainfall profile is not available, it is computed as a

percentage of the estimate of the volume expressed by $V_p^c + S_c$ which is here called V_{excess} (Eq. 7.19); this proportion is represented by α in Eq. 7.21. Note that $\hat{V}_{excess}$ (Eq. 7.20) depends on the magnitude of $\widehat{RO}_c^{daily}$ and S_o^{Cap} , i.e., the storage capacity of the inner river before the storm arrives. Thus, if S_o^{Cap} is small, $\hat{V}_{excess}$ could be small if $\widehat{RO}_c^{daily}$ is small. How large or small $\hat{V}_{excess}$ is depends on the value of $\widehat{RO}_c^{daily}$ and α, where the latter it is a value to be analysed in the framework.

As was mentioned above, the Type-1 pumping strategy can be applied by considering probabilistic and deterministic forecast information. The application and chronology of this pumping strategy when using these two types of information is explained as follows.

7.6.5.1 Type-1 pumping strategy based on deterministic forecast

The Type-1 pumping strategy under deterministic forecast information requires a forecast of the total daily runoff $\widehat{RO}_{daily}^c$ based on the deterministic forecast of a critical daily rainfall (R_{daily}^c) (a daily rainfall causing critical conditions in the polder), designed as $\hat{R}_{daily}^c$. By definition, the values of this latter variable are values greater than a daily rainfall threshold RT_{daily} (section 7.4), and, therefore, they are provided to the polder manager when a storm warning is issued. $\widehat{RO}_{daily}^c$ is computed here as:

$$\widehat{RO}_{daily}^c = 0.7\hat{R}_{daily}^c \qquad \text{Eq. 7.23}$$

where the value of 0.7 represents the average runoff coefficient of the polder system used in the conceptual model to compute the runoff rate RO (Eq. 7.1).

The chronology of the operation of the polder system under the Type-1 pumping strategy based on deterministic-24 forecasts can be summarized as follows.

- ✓ At midnight, a deterministic 24h-forecast of rainfall is generated and a warning decision is conducted based on Eq. 7.11. If a storm warning is issued, the deterministic forecast of the daily runoff that will cause critical conditions $\widehat{RO}_c^{daily}$ in the next 24 hours is delivered to the polder manager (Eq. 7.23). If a storm warning is not issued, only a reactive pumping action is conducted.

- ✓ If a storm warning is issued, the polder manager conducts the proactive action by pumping a volume of water equal to $\hat{S}_c = \alpha \hat{V}_{excess}$ with a pumping rate = q_{max}, where $\hat{V}_{excess}$ is computed as $\widehat{RO}_c^{daily} - S_o$ (Eq. 7.20).
- ✓ Then, the polder manager waits for the arrival of the storm. If the storm arrives before $\hat{S}_c$ has been pumped, the manager will continue with the proactive strategy and also implement the reactive strategy. Here, the pumping rate used is q_{max} until the target volume has been pumped.
- ✓ Finally, the polder manager completes the pumping strategy by conducting the reactive action once the storm arrives, which is represented by Eq. 7.15.

7.6.5.2 Type-1 pumping strategy based on probabilistic forecast

The Type-1 pumping strategy under probabilistic forecast information uses the expected value of the forecast of a daily rainfall R_{daily} to compute $\widehat{RO}_{daily}^{c}$.

$$\widehat{RO}_{daily}^{c} = 0.7\mathbf{E}\left(R_{daily}|\hat{R}_{daily}\right) \qquad \text{Eq. 7.24}$$

where $\mathbf{E}\left(R_{daily}|\hat{R}_{daily}\right)$ is the expected value of the conditional distribution of R_{daily} given $\hat{R}_{daily}$, i.e., $f\left(R_{daily}\middle|\hat{R}_{daily}\right)$, obtained from the joint probability of R_{daily} and $\hat{R}_{daily}$ (section 7.3.4).

The chronology of the operation of the polder system under the Type-1 pumping strategy based on probabilistic forecasts can be summarized as follows.

- ✓ At midnight, a probabilistic 24h-forecast of rainfall is generated, and a warning decision is conducted based on Eq. 7.12. If a storm warning is issued, the probabilistic-forecast based estimate of the daily runoff that will cause critical conditions $\widehat{RO}_c^{daily}$ in the next 24 hours is delivered to the polder manager (Eq. 7.24). If a storm warning is not issued, only a reactive pumping is conducted.
- ✓ If a storm warning was issued, the polder manager conducts the proactive action by pumping a volume of water equal to $\hat{S}_c = \alpha \hat{V}_{excess}$ with a pumping rate = q_{max}, where $\hat{V}_{excess}$ is computed as $\widehat{RO}_c^{daily} - S_o^{cap}$ (Eq. 7.20).
- ✓ Then, the polder manager waits for the arrival of the storm. If the storm arrives before $\hat{V}_{excess}$ has been pumped, the manager will continue with the proactive strategy, and also implement the reactive strategy. Here, the pumping rate used is q_{max} until the target volume has been pumped.
- ✓ Finally, the polder manager completes the pumping strategy by conducting the reactive action once the storm arrives, which is represented by *Eq. 7.15*.

7.7 Operation of the Shazhou polder system under different forecast scenarios

In this section, an explanation is provided of how the framework links the concepts explained above to build the algorithms to simulate the operation of the polder system under different forecast scenarios. In essence, the framework couples the RFG (section 7.3) with different warning decisions (section 7.5) and the conceptual model of the polder system (section 7.2.2) with different pumping strategies (section 7.6) according to the scenario to be analysed. Before introducing these algorithms, the metrics used to compare all the scenarios are explained.

7.7.1 Metrics to be analysed

A polder manager could use a warning system to reduce the time and magnitude of the waterlogging W by conducting a pumping strategy based on the forecasts (in this work 24-h forecasts). This pumping strategy can involve proactive and reactive actions with an inevitable pump operating cost, which should also be considered. There are, therefore, two criteria to be considered, waterlogging and pumping cost, and metrics to represent them are detailed below. Since the stochastic rainfall model on which the RFG is based represents the rainfall characteristics of July in Nanjing, when extreme events are most likely to occur, the metrics used in this work have been designed to represent the average values for that month. These values were computed by considering several replications of the operation of the polder system over the month of July. Each July replication is performed using continuous simulation, with the hourly simulation of the hydrology of the polder system carried out based on synthetic 'real' rainfall time series and the proactive decisions made using their corresponding (24hr) forecast values provided by the RFG.

7.7.1.1 Maximum Inundated area

The maximum inundated area (MIA) represents the maximum area inundated during each replication of the operation of the polder system during July. The inundated area (IA) is a value provided by the conceptual model of the polder system through simulating hourly runoff and calculating hourly inundation using Figure 7.4. MIA and its average value can be expressed by:

$$MIA = max(IA_1, IA_2, IA_3, \dots IA_j) \quad \text{Eq. 7.25}$$

$$\overline{MIA} = \frac{1}{n}\sum_{i=1}^{n} MIA_i \quad \text{Eq. 7.26}$$

where j is the total number of simulated hours in a simulated July, $\overline{MIA}$ is the average maximum inundated area in July, n represents the total number of July replications, and i is a simulated July. The $\overline{MIA}$ will be a function of the pumping strategy adapted, and the forecast information used.

7.7.1.2 Pumping costs

The pumping cost (C_p) represent the total cost of the pumping actions during July. The cost will depend on the given pumping strategy (and also the forecast information used) and can be split into proactive (C_{pro}) and reactive (C_{rea}) pumping costs. The average values of these pumping costs can be thus expressed as:

$$\bar{C}_{pro} = \frac{1}{n}\sum_{i=1}^{n} C_{pro}^{i} \quad \text{Eq. 7.27}$$

$$\bar{C}_{rea} = \frac{1}{n}\sum_{i=1}^{n} C_{rea}^{i} \quad \text{Eq. 7.28}$$

$$\bar{C}_p = \bar{C}_{pro} + \bar{C}_{rea} \quad \text{Eq. 7.29}$$

where C_{pro}^{i} and C_{rea}^{i} represent the total proactive and reactive pumping cost for a simulated July $i, \bar{C}_{pro}$ and $\bar{C}_{rea}$ represent the average proactive and reactive pumping costs during July, and $\bar{C}_p$ is the average total pumping cost during July.

The pumping costs were computed according to the assumed pumped tariff shown in Table 7.9. As one can see, the number of steps considered in the pumping rate is equal to 6. The number of steps in the pumping rate has been assumed considering that the number of pumps at each pumping station in the Shazou polder, which is also 6.

Table 7.9: Assumed pumping tariff to compute the pumping costs

q [mm h^{-1}]	Pumping tariff [Units]
1.6	15
3.2	30
4.8	45
6.4	60
8	75
q_{max} =9.62	100

7.7.1.3 Duration of waterlogging

The duration of waterlogging (D_w) represents the number of hours of waterlogging during a simulated July. The average value is given by:

$$\overline{D}_w = \frac{1}{n}\sum_{i=1}^{n} D_w^i \qquad \text{Eq. 7.30}$$

Where D_w^i represents the number of hours of waterlogging in the polder system during a simulated July i, and $\overline{D}_w$ is its average.

7.7.2 The no forecast scenario

The no forecast scenario will be one of the benchmark cases in the framework. It represents the current pump operation procedure in the polder system, which is based on reactive pumping actions. It will be represented by using the algorithm we built to represent the water balance in the polder (section 7.2.2) with the pumping strategy described in section 7.6.3. The algorithm designed for its analysis is detailed below.

- ✓ **Step 1:** Obtain a continuous set of 'observed' hourly synthetic rainfall time series for July from the RFG.
- ✓ **Step 2:** Run the analysis in no-warning mode using the conceptual model of the polder system with the pumping strategy described in 7.6.3.
- ✓ **Step 3:** For all the simulated Julys, compute the metrics explained in section 7.7.1

7.7.3 The perfect forecast scenario

The perfect forecast scenario will be the second benchmark scenario in the framework. It is an idealized scenario in which there is no uncertainty for the target

variables in the proactive pumping strategy, i.e., it is assumed that the hourly rainfall values are known, and 24hr forecasts are not used. It will be represented by using the algorithm built to characterize the water balance in the polder (section 7.2.2) with the pumping strategy described in section 7.6.4. The algorithm designed for its analysis is detailed below.

- ✓ **Step 1:** Obtain a continuous set of 'observed' hourly synthetic rainfall time series for July from the RFG.
- ✓ **Step 2:** Run the analysis in no-warning mode using the conceptual model of the polder system with the pumping strategy described in 7.6.3.
- ✓ **Step 3:** For all the simulated Julys, compute the metrics explained in section 7.7.1

7.7.4 The deterministic-forecast scenario

The deterministic forecast scenario is one of the imperfect forecast scenarios considered in the framework. It uses deterministic 24hr-forecasts to represent both the warning decision and the proactive pumping strategy. It will be represented by using the decision rule that describes the deterministic warning decision (section 7.5.1) and the algorithm built to characterize the water balance in the polder (section 7.2.2) with the Type-1 pumping strategy described in 7.6.5.1. The algorithm designed for its analysis is detailed below.

- ✓ **Step 1:** Obtain a continuous set of 'observed' and 'forecast' hourly synthetic rainfall time series for July from the RFG. Aggregate the 'forecast' hourly rainfall to obtain 24-h forecasts.
- ✓ **Step 2:** For each July replication, run the analysis in deterministic-forecast mode, i.e., by using the conceptual model of the polder system, for each day:
  - Compute $\hat{R}_{daily}$ and simulate the deterministic warning decision based on Eq. 7.11
  - Run the conceptual model of the polder system in concert with the Type-1 pumping strategy described in 7.6.5.1.
- ✓ **Step 3:** For all simulated Julys, compute the metrics explained in section 7.7.1

7.7.5 The probabilistic-forecast scenario

The probabilistic forecast scenario is the other imperfect forecast scenario considered in the framework. It uses 24-hr probabilistic forecasts to represent both the warning decision and the proactive pumping strategy. It will be represented by using the decision rule that describes the probabilistic warning decision (section 7.5.2) and the algorithm that was built to characterize the water balance in the polder

(section 7.2.2) with the chronology of the Type-1 pumping strategy described in 7.6.5.2. The algorithm designed for its analysis is detailed below.

- ✓ **Step 1:** Obtain a continuous set of 'observed' hourly synthetic rainfall time series for July from the RFG. Aggregate the 'forecast' hourly rainfall to obtain 24-h forecasts.
- ✓ **Step 2:** For each July replication, run the analysis in probabilistic-forecast mode, i.e., by using the conceptual model of the polder system, for each day:
    - Compute $\hat{R}_{daily}$ and its associated $f\left(R_{daily}\middle|\hat{R}_{daly}\right)$ from the joint distribution of R_{daily} and $\hat{R}_{daly}$ (section 7.3.4.2), and compute PE based on RT_{daily}.
    - Simulate the probabilistic warning decision based on Eq. 7.12.

  Step 3: Run the conceptual model of the polder system in concert with the Type-1 pumping strategy described in 7.6.5.2.
- ✓ **Step 4:** For all simulated Julys, , compute the metrics explained in section 7.7.1

7.8 Scenario simulation for a single storm

The prior section explains the framework algorithms used to compare the operation of the polder system under different forecast scenarios. The outputs of these algorithms are the metrics explained in section 7.7.1. If one shows only the results of the framework based on these algorithms, one cannot fully appreciate how the several pumping strategies considered in the analysis work during a critical storm. This section, therefore, shows an example of the simulation of the polder system under all scenarios during the same observed storm causing critical conditions. This analysis will give us a good insight into how the pumping criteria adopted for the different strategies are reflected when trying to mitigate a critical condition situation. Figure 7.17 and Figure 7.18 show this example, and the results for each scenario are explained below. In this case, the initial condition for all scenarios matches. Note on the first day, the storm magnitude is insufficient to trigger a warning in the proactive scenarios, and reactive pumping is performed in all cases. During day 2, a storm triggers proactive action in all cases, except for the reactive strategy. A general description of the scenarios is given in Table 7.10.

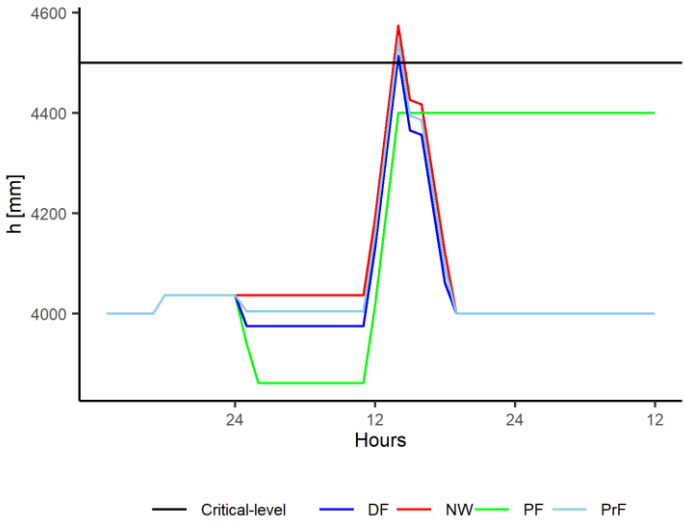


Figure 7.17: Example of the simulation of the operation of the polder system (water level) during an observed storm causing critical conditions for all scenarios

In the figure, DF: deterministic forecast scenario; NW: no warning scenario; PF: perfect forecast; PrF: probabilistic forecast scenario. This figure shows how the pumping strategies considered in each scenario work during an observed storm causing critical conditions. Table 7.10 gives a general description of these scenarios. In this case, the initial condition for all scenarios matches. The normal water is 4000 mm, and $\alpha =$ 0.05 and 0.025 for the deterministic and probabilistic scenario, respectively. h_{ref} = 4400 for the perfect forecast scenario.

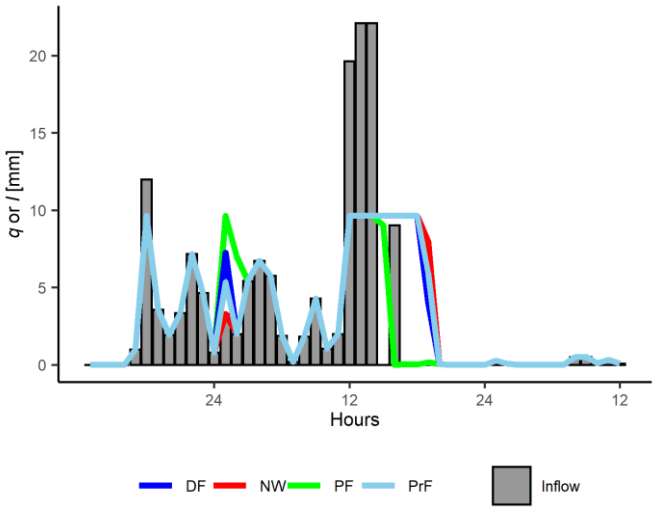


Figure 7.18: Example of the simulation of the operation of the polder system (pumping rate and Inflow) during an observed storm causing critical conditions for all scenarios

In the figure, DF: deterministic forecast scenario; NW: no warning scenario; PF: perfect forecast; PrF: probabilistic forecast scenario. This figure shows the inflow I and pumping rate q of the example shown in Figure 7.17. The maximum inflow I is equal to the r, the capacity of the municipal pipe network, and the maximum q is equal to the pumping capacity of the polder q_{max}.

- ✓ ***Non-warning scenario***: This scenario represents the current pump operation for the Shazou polder, which is a reactive pumping strategy (section 7.6.3). When the water level starts to rise, the inflow I is greater than q_{max}, and the polder manager will pump the water with a pumping rate equal to q_{max}. After the critical condition situation, the water level of the

inner rivers is dropped to the normal water level h_n with a pumping rate equal to q_{max}.

- ✓ ***Perfect forecast scenario***: Under this scenario, the polder manager has perfect knowledge about S_c, and the value of V_p^{before} adopted in Eq. 7.14 is $S_c + (h_c - h_{ref})$(Eq. 7.16). Thus, the water level is dropped before the storm arrives, and the maximum water level matches with h_{ref} that is here assumed to be 4400 mm.
- ✓ ***Deterministic-forecast scenario***: Under this scenario, the value of V_p^{before} adopted in Eq. 7.14 is $\hat{S}_c$ (Eq. 7.21) estimated based on the deterministic forecast of daily rainfall. Figure 7.17 shows the results when adopting a value of α=0.05. As can be seen, for this storm, a warning was issued, and the value adopted for α is not enough to avoid the critical condition. Therefore, after the critical condition situation, the water level is dropped to the normal water level h_n with a pumping rate equal to q_{max}.
- ✓ ***Probabilistic-forecast scenario***: Under this scenario, the value of V_p^{before} adopted in Eq. 7.14 is $\hat{S}_c$ (Eq. 7.21) estimated from $\hat{R}_{daily}^c$ computed based on the expected value of forecast of daily rainfall (Eq. 7.24). Figure 7.17 shows the results when adopting a value of α=0.025. As can be seen, for this storm, a warning was issued, and the value adopted for α is not enough to avoid the critical condition. Therefore, after the critical condition situation, the water level is dropped to the normal water level h_n with a pmping rate equal to q_{max}

Table 7.10 General description of the scenarios considered in the framework

Scenario	Warning decision	Pumping strategy
The no-warning scenario (NW)	No	Reactive pumping strategy (section 7.6.3).
The perfect forecast scenario (PF- add for each)	No	It assumes perfect knowledge of the target variables S_c and V_p^c (section 7.6.4)
The deterministic forecast scenario	Based on the deterministic forecast of the daily rainfall $\hat{R}_{daily}$ and the daily rainfall threshold RT_{daily}	✓ Based on the forecast of the daily rainfall $\hat{R}_{daily}$ ✓ The critical daily runoff $\widehat{RO}_{daily}^c$ is computed as $\hat{R}_{daily}$*0.7. ✓ The proactive pumping is a function of the parameter α (proactive pumping factor).
The probabilistic-forecast scenario	Based on the probabilistic threshold PT and the exceedance probability PE of RT_{daily}	✓ Based on the expected value of R_{daily}, i.e., $\mathbf{E}(R_{daily}\|\hat{R}_{daily})$. ✓ $\widehat{RO}_c^{daily}$ is computed as $\mathbf{E}(R_{daily}\|\hat{R}_{daily})$*0.7 ✓ The proactive pumping is a function of the parameter α (proactive pumping factor).

7.9 Tradeoff between inundated area and pumping cost

In this section, the framework is used to compare the performance of the drainage system in the Shazhou polder in July under different forecast scenarios. As the title of this section implies, the results show a tradeoff between the metrics that represent the inundated area and the pumping costs. Two parameters will be analysed in the framework through the scenarios: α and PT. The aim of the analysis of these parameters in each scenario is described in Table 7.11. The results of the deterministic and probabilistic scenarios will be first shown separately, and, then, all the scenarios will be analysed together. The imperfect forecast scenarios will be compared with the benchmark scenarios, i.e., the no warning and the perfect forecast scenarios. The analysis considered Julys with at least one rainfall event that could potentially produce a significant runoff event in the polder system (daily rainfalls > 50 mm). After this filter, the sample size was reduced to 8730 Julys. The results are provided below.

Table 7.11 Parameters to be analysed in the framework

Scenario	Parameter	Range	Aim
The deterministic scenario	α	0-0.5	Analyse the tradeoff between pumping cost $\bar{C}_p$ and inundated area $\overline{MIA}$.
The probabilistic scenario	α, PT	α=0-0.5 PT=0.1-1	Analyse the best solutions (Pareto curve) in terms of $\bar{C}_p$ and $\overline{MIA}$.

7.9.1 Analysis of the tradeoff for the deterministic-forecast scenario

The results of the deterministic forecast scenario are provided in Figure 7.19, showing the tradeoff between the pumping costs and the inundated area. Note that the average proactive pumping costs $\bar{C}_{pro}$ represents between 1 and 3 % of total average costs $\bar{C}_p$ due to the latter considering not only the reactive and proactive pumping for a critical daily rainfall but also the reactive pumping cost of all daily rainfalls in the month.

The results of the benchmark cases are also plotted in Figure 7.19. The no warning condition defines the worst scenario in the analysis due to no proactive action being taken in the polder. In contrast, the perfect forecast scenario represents the best scenario due to perfect knowledge of the target variables in the pumping strategy.

Note that the pumping costs of this scenario are slightly lower than the no warning scenario. That occurs because the number of critical condition situations decreases significantly in this scenario, and thereby the number of times the polder manager drops the water level of the inner rivers with a pumping rate equal to q_{max}. This is reflected in the reduction of the reactive pumping costs, which also reduce the total pumping costs. Thus, any strategy of the deterministic forecast scenario (and any imperfect forecast strategy) cannot overcome these results; they are explained as follows.

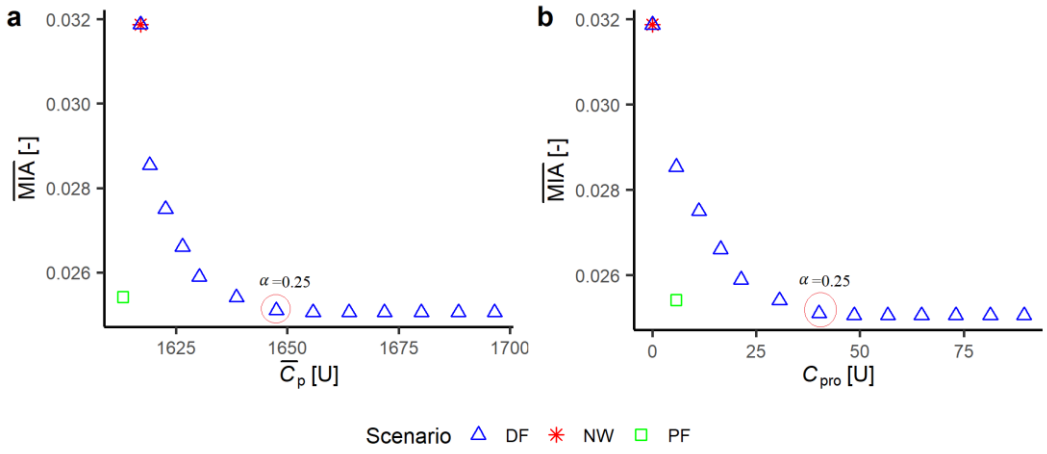


Figure 7.19: Tradeoff between pumping costs and inundated area for the deterministic forecast scenario and comparison with the two benchmark scenarios

In the figure, DF: deterministic forecast; NW: no warning scenario; PF: perfect forecast. This figure shows the tradeoff between $\bar{C}_p$ and $\overline{MIA}$ (a) and $\bar{C}_{pro}$ and $\overline{MIA}$ (b) for the deterministic forecast scenario. The values of α considered were the following: 0, 0.025,0.05,0.075,0.10,0.15,0.2,0.25,0.30,0.35,0.40,0.45,0.5. It also shows the values of the two benchmark scenarios. In these figures, the values of $\overline{MIA}$ decrease as the values of α increase.

The worst strategy of the deterministic forecast scenario is when $\alpha = 0$ (the highest triangle on the plots), i.e., when the volume of water pumped before a critical storm is zero, which can be considered as a reactive pumping strategy. Therefore, as one can expect, this strategy matches the results of the no warning scenario. As α increases, the inundated area decreases, and the pumping costs increase. However, there is a point (α=0.25) where the inundated area stops decreasing, and it stays constant, which means that, after this point, critical storms cannot be avoided. That occurs because, after this point, most of the remaining critical storms to be avoided are those whose runoff starts at midnight (or close to this time) and whose inflow rate overcomes the pumping capacity of the polder q_{max}. These storms are also a problem for the perfect forecast strategy. Under this condition, the polder manager does not have the required response capacity for the critical storm, and he/she can

only use a pumping rate equal to q_{max}, whereas the water level of the inner rivers rises until a critical condition situation occurs. Thus, after $\alpha \geq 0.25$, the values of $\overline{MIA}$ are associated with waterlogging caused by these storms and by storms whose runoff overcomes the capacity of the drainage system.

Another important variable to be analysed is D_w, i.e., the number of hours of waterlogging in the polder system during July. The relation of this variable with the pumping cost is shown in Annex 2. They have the same shape as the figures explained above; therefore, they do not need any further explanation. To have a measure of the magnitude of this variable, the 99-percentile of values of D_w greater than 1 hour was computed (D_w^{99}). The results are shown in Figure 7.20. As one can see, the behaviour of the scenarios is the same as the one shown in the prior figures, where the no warning and perfect forecast scenario represents the worst and the best ones, respectively, with the deterministic forecast located in the middle.

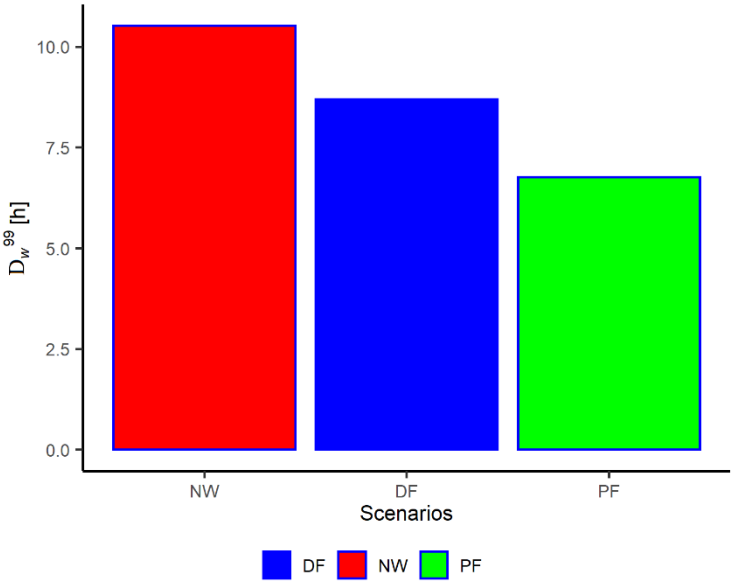


Figure 7.20: Values of D_w^{99} for the deterministic strategy and the benchmark cases
In the figure, DF: deterministic forecast; NW: no warning scenario; PF: perfect forecast. The values of α assumed for the deterministic pumping strategy was 0.15

7.9.2 Analysis of the tradeoff for the probabilistic forecast scenario

In this section, the analysis of the probabilistic scenario is conducted. It aims to find the best solutions for the set of parameters (PT,α) considered when simulating the polder system under this scenario. The approach consisted of simulating the polder system by considering different values of PT in the warning decision for each value of α in the Type-1-pumping strategy based on probabilistic information (section

7.6.5.2). Then, the plot of $\bar{C}_p$ vs $\overline{MIA}$ was used to define a set of "best" points as the Pareto front. The results are shown in Figure 7.21, and the values of the best strategies are shown in Appendix B.

The best points showed different combinations of probability PT and α, with several values of PT sharing the same value of α. However, if the polder manager chooses a value of α, it would be preferable to have just one value of PT that performs better than the deterministic scenario. This analysis was performed by plotting the best Pareto solutions associated with the same α with different values of PT and comparing them with the deterministic result associated with the same value of α. Then, the point that overcomes the deterministic result and is closest to the perfect forecast scenario was chosen. Figure 7.22 shows one example of this analysis for one of these sets of best solutions. After this analysis, the set of best solutions were reduced to seven pairs with unique values of α with different PT values. The values of this set of best probabilistic solutions are shown in Table 7.12.

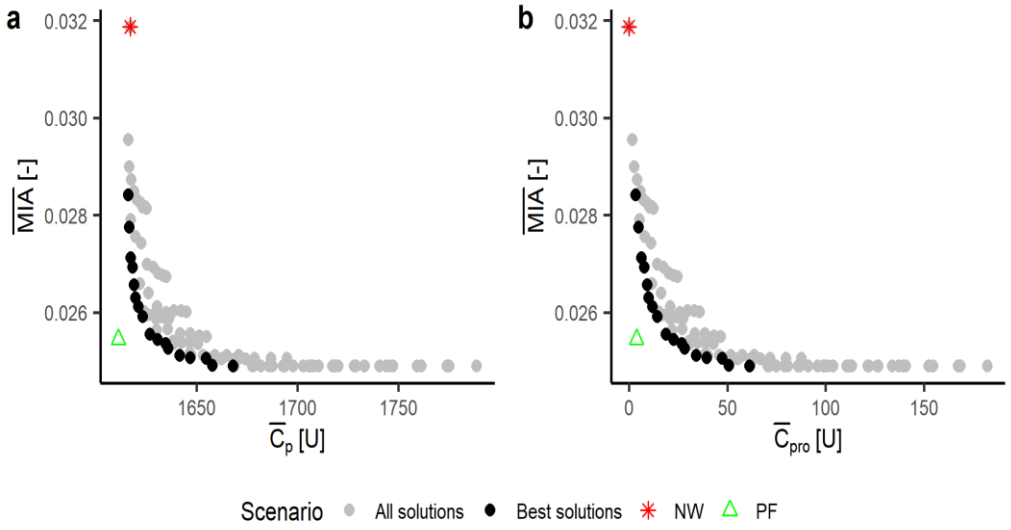


Figure 7.21: Tradeoff between pumping costs and inundated area for the probabilistic forecast scenario and comparison with the two benchmark scenarios

In the figure, NW: no warning scenario; PF: perfect forecast. This figure shows the tradeoff between $\bar{C}_p$ and $\overline{MIA}$ (a) and $\bar{C}_{pro}$ and $\overline{MIA}$ (b) for the probabilistic forecast scenario. when assuming different values of α and PT. The values of α considered were the following: 0, 0.025,0.05,0.075,0.10,0.15,0.2,0.25,0.30,0.35,0.40,0.45,0.5., whereas the values of PT were 0.1,0.2,0.3,0.4,0.5,0.6,0.7,0.8,0.9 It also shows the values of the two benchmark scenarios.

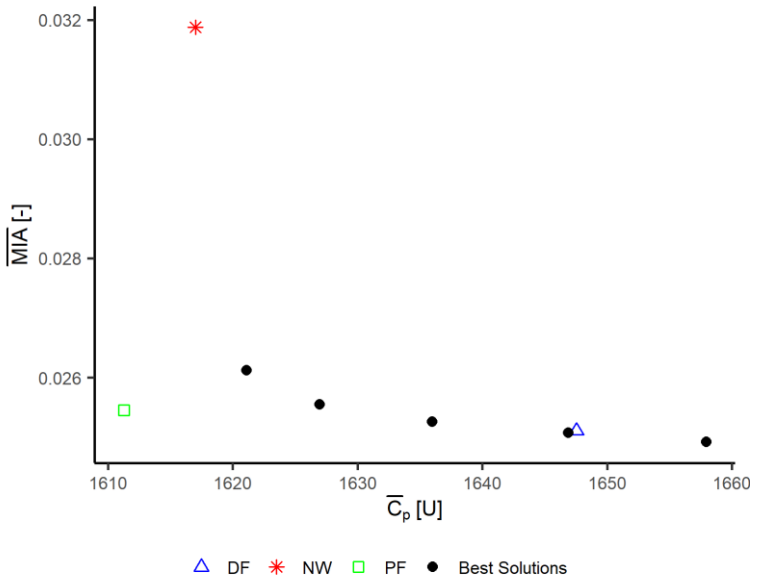


Figure 7.22: Example of when one should choose among best probability solutions with the same value of α based on the deterministic results and perfect forecast scenario

In the figure, DF: Deterministic forecast; NW: no warning scenario; PF: perfect forecast. This figure shows one of the sets of best solutions shown in Figure 7.21 with the same values of $\alpha = 0.25$ and different *PT* values. Only one pair of this set of values was chosen. The criterion adopted was to choose the pair whose point overcomes the deterministic result and is closest to the perfect forecast scenario.

Table 7.12. The final set of the parameters for the probabilistic forecast scenario

Set	α	PT
1	0.05	0.9
2	0.075	0.8
3	0.1	0.8
4	0.15	0.9
5	0.2	0.7
6	0.25	0.7
7	0.3	0.8

7.9.3 Analysis of the tradeoff for all scenarios

The final analysis in the framework consisted of comparing all scenarios together. This section primarily focuses on comparing the results of the deterministic and probabilistic forecast scenarios. Figure 7.23 and Figure 7.24 show the results for all scenarios analysed above in terms of pumping costs, inundated area and D_w^{99} respectively. As one would expect, the results of the probabilistic scenario are better than the deterministic one, i.e., the points in the plot are closer to the perfect information point. Thus, one can conclude that the probabilistic strategy is the best

one of the imperfect forecast scenarios; but understandably, it cannot be as good the perfect forecast strategy.

To summarise, the framework shows that the non-warning scenario defines the highest flood impact which is expressed by the highest values of $\overline{MIA}$ and $\bar{D}_w$ across all scenarios The perfect-forecast scenario is clearly best ; any strategy based on imperfect information cannot match the results of this scenario. The strategy used in the imperfect forecast scenario (Type-1 pumping strategy, section 7.6.5) demonstrates the benefits of the warning system when using imperfect forecasts of the daily rainfall. The benefits can be valued when comparing $\bar{C}_p$ with $\overline{MIA}$ or $\bar{D}_w$. The results of the probabilistic forecast scenario are better than those for the deterministic-forecast scenario. One can see clearly that when using probabilistic information, the benefits of the warning system increase.

The perfect forecast point in Figure 7.23 and Figure 7.24 provide an important benchmark against which to judge how improvements to the imperfect forecasts could be made. e.g., by improving the rainfall forecasts. The correlation between the observed and forecasts 24-h rainfalls was 0.90 (Table 7.5), which roughly corresponds to an explained variance of R_{daily}. This gives an idea of how useful such forecasts would be in the management of the polder and the quality of 24-h forecasts needed to achieve the results demonstrated.

Depending on how the polder manager views the tradeoff between cost and inundated area, a value of α can be chosen to reflect the tradeoff. If he/she wishes to minimize inundation and cost, then the best value would be that corresponding to the point where the plots flatten out, i.e., no reduction in inundated area can be achieved by incurring additional pumping costs.

For the framework simulation experiments conducted here, the focus has been on demonstrating the value of rainfall forecasts in polder management. A perfect model of the polder system has been assumed to do this. For an operational system, a fully calibrated model would be needed, and the uncertainty associated with polder model predictions would have to be taken into account.

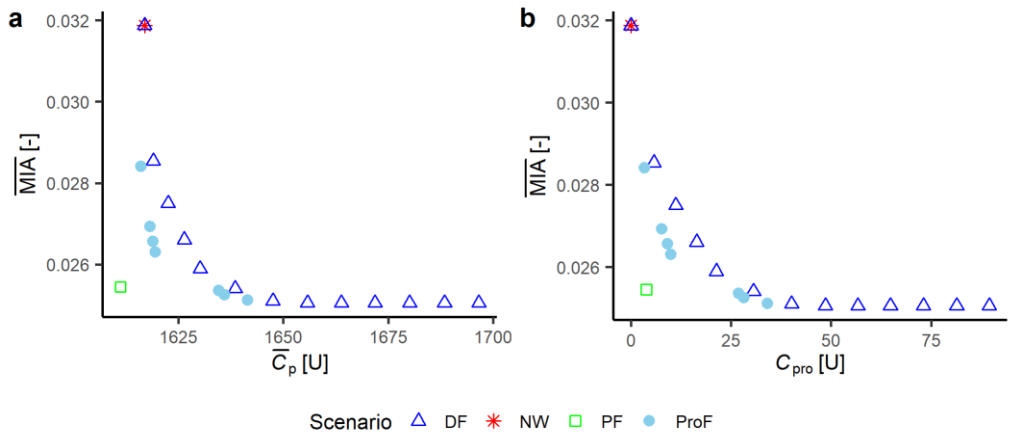


Figure 7.23: Analysis of the tradeoff for the deterministic and probabilistic scenarios and comparison with the benchmark scenarios

In the figure, DF: deterministic forecast; NW: no warning scenario; PF: perfect forecast; ProF: Probabilistic forecast. This figure compares the results of the deterministic and probabilistic forecast scenarios explained above.

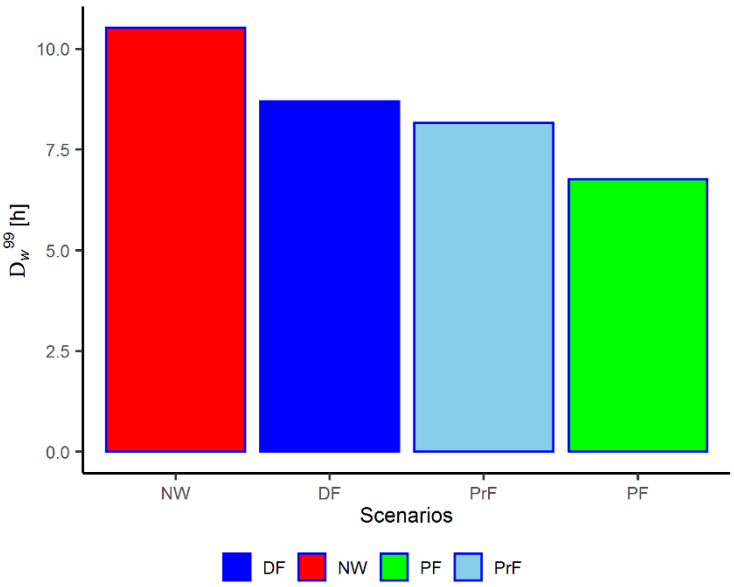


Figure 7.24: Values of D_w^{99} for all scenarios

In the figure, DF: deterministic forecast; NW: no warning scenario; PF: perfect forecast, PrF: probabilistic forecast. The values of α assumed for the deterministic and probabilistic pumping strategy was 0.15

7.10 Main findings

In this Chapter, it has been shown that the warning and pumping strategies adopted in a FEWS operating for flood-prone polder were important factors controlling the performance of the system. Through simulation experiments, it was observed that deterministic-forecasts-based strategies produced performance measures that can be enhanced by probabilistic-forecasts-based strategies. The results of the research

showed a trade-off between the average pumping costs $\bar{C}_p$ and the measures defining the performance of the FEWS ($\overline{MIA}$ and $\bar{d}_w$).

Given the modelling approach adopted to obtain the results mentioned above, this Chapter also provides new knowledge in the simulation and design of this type FEWSs. In this topic, one can say that the architecture of the RFG is novel; it gives a criterion to represent forecast and observed rainfall for the analysis of forecast-rainfall-based FEWSs. The results showed that R_{daily} and $\hat{R}_{daily}$ fitted well to the three-parameter Gamma distribution, suggesting, thus, that the seven-parameter BGM was a suitable model to represent the bivariate relationship of these variates. Furthermore, it was observed that the bivariate simulation of bivariate Gamma variates through the Gaussian copula is not versatile since this bivariate distribution does not hold an analytical expression that links its dependence structure with the correlation coefficient of the Gaussian copula $\rho_{\eta\hat{\eta}}$.

A rainfall threshold RT curve was used as a tool for issuing flood warnings. The methodology used to build that curve and the (deterministic and probabilistic) decision rules used to simulate warning decisions are novel. They represent a significant contribution in this field. Furthermore, the pumping strategies and the resulting waterlogging W in the polder area after they have been done were simulated based on the simple stylized model. This model has not been applied before in a flood warning context and was linked with an impact curve to represent the response and impact component (RIC) in the integrated framework. This model was not meant to be a precisely calibrated model but an approximate representation of reality.

Chapter 8. Discussion, conclusions and recommendations for future work

This thesis evaluates the performance of a simulated flood early warning system (FEWS) monitoring and warning a generic fluvial and fluvial flood-prone area as a function of several controlling factors. The evaluation is carried out within a novel generic framework where FEWS performance depends on the effectiveness of each of the components of a FEWS and on component linkages. The research contributions have provided a new understanding of the factors controlling the performances of FEWSs that can support their design. The knowledge gained can, thus, be used to i) understand the main factors controlling the performance of an operational FEWS and identify which of them can be positively modified to improve the system's performance, and ii) to design and/or simulate each of these FEWSs.

A real-world application of this versatile framework has been demonstrated through a case study of a flood-prone polder area in which a FEWS has been designed and simulated and shown to have significant potential operational benefits in managing waterlogging while having the opportunity to also reduce pumping costs.

In the next subsection, a general discussion of the main results obtained in this thesis is given. Then, the general conclusions are highlighted, and, finally, future work is suggested.

8.1 General discussion

The general discussion has been split into topics that are related to the research objectives. This subsection first discusses the main results according to these topics and ends by emphasizing this thesis's main contributions.

8.1.1 The design of the generic fluvial framework

An integrated and versatile generic simulation framework has been designed, which has achieved Objective 1. Moreover, this framework has enabled the overall aim of the thesis to be achieved by linking forecast model outputs with warning decisions and the associated responses and flood impacts, thus allowing the main factors influencing the performance of the integrated system to be evaluated. The

framework's versatility is due to the ability to sample extensively synthetic forecast and observed data using appropriate analytical expressions to represent their properties which are representative of real catchment data. The warning and response/impact processes are evaluated in terms of reliability and economic effectiveness. This versatility has allowed situations to be explored that are impossible with real-world modelling due to a lack of data or limited computational resources.

The framework is versatile because several bivariate distributions and generation algorithms (Ch4), warning performance measures and decision rules (Ch5 and Ch7), and impact and response functions, based on damage and cost (Ch6 and Ch7), can be employed. Moreover, deterministic and probabilistic forecasts can be simulated, and the predictive uncertainty (PU) of the latter can be used to optimize flood warning decision rules. The new framework avoids the restrictive assumptions made in previous frameworks.

8.1.2 The building of the MCFG

In recent years, a lot of effort has been put into the design of flood forecasting techniques to quantify or reduce PU in operational FEWS. Thus, in the literature, one can find many studies suggesting forecasting models for these purposes (Krzysztofowicz and Kelly, 2000; Coccia and Todini, 2011; Hapuarachchi *et al.*, 2011; Jain *et al.*, 2018). These flood forecasting techniques are complex and are rarely used in frameworks that include the completed integrated system (the FEWS). In this context, the benefits of quantifying or reducing PU have been poorly studied in an integrated framework. To overcome this issue, this research has used a Monte Carlo flood and forecast generator (MCFG), which, in essence, simulates the forecasts generated by these forecasting models with their associated observations, and most importantly, enables the PU of the forecast to be quantified.

In both the simulated generic fluvial case and flood-prone-polder system case study, the MCFG has used the concept of PU to represent/quantify the uncertainty about the future value, conditional on a single-value forecast of the predictand. Even though the aim of the MCFG for each case was the same, i.e., generate potential scenarios of flooding in an at-risk community, the predictand and predictor considered and the architecture of the MCFG was different in each of them. In the

simulated generic fluvial case, a potential flood in the at-risk community was defined by observed peak water levels (y). Thus, y was assumed to be the predictand, and its forecast ($\hat{y}$), the predictor. In the simulated flood-prone-polder system case study, a flood in the polder area was predicted based on a 24-hour forecast of significant rainfall. Thus, here, the significant daily rainfall (R_{daily}) was assumed to be the predictand, and its forecast ($\hat{R}_{daily}$), the predictor. The architecture of the MCFG is novel; it allows PU to be expressed as a function of forecasting lead time in the simulated generic fluvial case and, in the simulated flood-prone polder system case study, gives a criterion that allows rainfall PU to be expressed for the analysis of rainfall forecast-based FEWSs.

A bivariate parametric model was used within the MCFG to do a bivariate simulation of pairs (y, $\hat{y}$) with its associated conditional distribution $f(y|\hat{y})$ in the simulated generic fluvial case, and to estimate PU in terms of $f\left(R_{daily}\middle|\hat{R}_{daily}\right)$ in the simulated flood-prone polder system case study. Since a FEWS should serve its function during every potential flood, not just the largest potential flood in a year or month, the marginal distributions of these bivariate parametric models were represented by peak-over-threshold (POT) models. The bivariate parametric model used in each case was based on univariate analysis of observed sample data. In the simulated generic fluvial case, the sample data consisted of values of y for four gauging stations located close to or at a floodplain in northern England (Figure 4.4), and of the values of $\hat{y}$ of one of these stations. The latter values were represented by simulated discharges from an existing calibrated model, which were then converted into water levels using the at-site gauge rating curve. The results relating to y for each gauging station showed that at least two of the following univariate parametric distributions fit this variate well: the two-parameter Exponential distribution, the three-parameter Lognormal distribution, and the three-parameter Gamma distribution, whereas the two-parameter exponential distribution fitted the values of $\hat{y}$ well (a goodness of fit test could only be applied in this case). For the station where values of y and $\hat{y}$ were both available, the same probability distribution (the two-parameters exponential) fitted both sample values well. This result allowed the assumption to be made that the probability distribution type of $\hat{y}$ is the same as that of y which can be used when values of the former variate are not available, which is very common in operational FEWSs. Based on this assumption, and the results of the univariate analysis, the five-parameter bivariate Exponential

distribution (BED), the seven-parameters bivariate Lognormal distribution (BLND), and the seven-parameter bivariate Gamma distribution (BGM) were considered to be suitable models to represent the pairs (y, $\hat{y}$). Since the simulated generic fluvial case was based on a virtual floodplain and did not represent any of the floodplains for which the frequency analysis was carried out, the seven-parameters BLND was chosen for simulating the pairs (y, $\hat{y}$) used in the sensitivity experiments carried out in Chapters 5 and 6.

The results of the univariate analysis of y and $\hat{y}$ match with other POT-frequency analysis studies of extreme hydrological variables (Choulakian *et al.*, 1990; Bezak *et al.*, 2014) that have been carried out. Since none of these works and other related works (Claps and Laio, 2003; Bogner *et al.*, 2012) have done a frequency analysis of y and $\hat{y}$, there is no formal criterion to define the pre-assumed threshold above which all peak water levels occur. In the univariate analysis of y and $\hat{y}$, a stage level (y_b) must be set neither so high that only few floods are considered in the hydrologic frequency analysis, i.e., it should be lower than the flooding threshold (y_T), nor so low that too many peaks are considered which are not relevant in the analysis. The value of y_b considered in this work was a stage level that has been set by the Environment Agency for each gauging station used for flood warning purposes in England. When the river's water level reaches that threshold, a minor flood is possible in the floodplain. In this context, this research gives, at least for England, a formal criterion to define y_b which is a very sensitive parameter in the POT frequency analysis.

The univariate analysis in the simulated flood-prone polder system case study was based on a sample of synthetic data obtained from the stochastic rainfall field generator (RainSim V3) embedded in the MCFG; the RainSim model was calibrated using hourly and daily data records for the case study site. Pairs of correlated daily rainfall totals (R_{daily},$\hat{R}_{daily}$) greater than 50 mm were obtained from the generated sample to include in the analysis; daily rainfall above this threshold could potentially produce significant runoff in the polder system. The results showed that the three-parameter univariate Gamma distribution fitted well both R_{daily} and $\hat{R}_{daily}$, suggesting, thus, that the seven-parameter BGM was a suitable model to represent the bivariate relationship of these variates.

A Gaussian copula-based algorithm was used within the MCFG to build $f(y|\hat{y})$ and $f\left(R_{daily}|\hat{R}_{daily}\right)$. To define the value of the correlation coefficient of the Gaussian copula ($\rho_{\eta\hat{\eta}}$) associated with the correlation coefficient in the real space, which controls the PU, different techniques were used in each case. In the simulated generic fluvial case, an analytical expression that defines the relationship between the correlation coefficient of the (real) Lognormal space (ρ_{LN}) and $\rho_{\eta\hat{\eta}}$ in the Normal space was used. In the flood-prone polder system case study, an equation that defines the relationship between the correlation coefficient of the (real) Gamma space (ρ_G) and $\rho_{\eta\hat{\eta}}$ in the transformed space for the assumed marginal distribution of R_{daily} and $\hat{R}_{daily}$ was built via Monte Carlo (MC) simulation. These results suggested that, when using a parametric bivariate model to represent the predictand and predictor based on the approach described in this research, one should try to use a model that incorporates an analytical expression that links its dependence structure with $\rho_{\eta\hat{\eta}}$. By doing that, the MCFG and the integrated framework (Figure 3.1) gain versatility. This was a significant reason for choosing the bivariate Lognormal for the simulation experiments in Chapters 5 and 6.

8.1.3 Warning decisions and warning criteria

In the analysis of FEWSs, the flood warning decision issue has not received the same attention as the prediction problem. The flood warning decision process is complex, as several factors are involved (Arnal *et al.*, 2020). This process is conducted according to the warning criterion adopted in the FEWS. A common practice is to simplify this process through decision rules representing rational decisions (Verkade and Werner, 2011; Girons Lopez *et al.*, 2017; Bischiniotis *et al.*, 2019). This research used decision rules for representing warning decisions based on deterministic and probabilistic forecasts, which were set based on the warning criterion adopted in each simulated case.

In the simulated generic fluvial case, FEWSs with and without real-time flood maps were considered, and deterministic and probabilistic warning decision rules were set for each of them; they are explained in Chapter 3 (Figure 3.3). The probabilistic-threshold-based decision rule (PTDR) used to simulate FEWSs without real-time-flood maps have been widely used in several research works (Coccia and Todini, 2011; Verkade and Werner, 2011; Girons Lopez *et al.*, 2017); however, the

probabilistic-based decision rule (PDR) used to simulate FEWSs with real-time flood maps is introduced in this thesis, and is novel. This rule uses a warning level ($\hat{y}_w$) derived from $f(y|\hat{y})$ to make the warning decision, where $\hat{y}_w$ is defined by a probabilistic threshold (PT), which has to be optimised based on a pre-defined criterion. This level defines the number of houses that should be warned in the floodplain using a surrogate function that represents the number of houses to be warned as a function of $\hat{y}_w$. The PDR was used because FEWSs based on real-time-flood maps must use the flood map corresponding to $\hat{y}_w$ which in turn must be defined by optimizing a probabilistic threshold. Since PTDR is based on a probability value that determines whether or not to warn ALL houses in the floodplain, it cannot be used to simulate this type of FEWS. That can be considered a drawback of the PTDR if it is used to simulate a FEWS as the impact or the cost of the warning response (C_w) cannot be estimated as a function of the flood magnitude's prediction (Verkade and Werner, 2011). Moreover, it is proposed here that performance measures such as POD and FAR should be computed based on the house warned and flooded/not flooded. This criterion, called the floodplain property-based criterion (FPC) in Ch5, has been shown to be a more demanding criterion as it differentiates between warned and flooded houses when a flood is preceded by a warning (hit), unlike the flooding threshold-based criterion (FTC) that was also considered in Ch5 and other research works (Verkade and Werner, 2011; Pappenberger *et al.*, 2015; Bischiniotis *et al.*, 2019), which assumes that all houses are flooded if it is hit (Figure 5.4).

In the simulated flood-prone polder system case study, a rainfall threshold (RT) curve was used as a tool for simulating flood warnings. This curve was built considering the uncertainty of the daily rainfall profile and the initial condition of the water level of the inner rivers when the forecast is issued. The RT curve and two decision rules were then used to simulate flood warning decisions under the deterministic and probabilistic forecast scenarios. There are few works that have considered the uncertainty of rainfall characteristics when analysing rainfall thresholds for flood warning systems (Wu *et al.*, 2015), and, to the best of the author's knowledge, a methodology to quantify rainfall thresholds for flood warning purposes in a polder system has not been published yet.

8.1.4 The response and impact simulation

The study of the response and impact component (RIC) for the analysis of FEWSs through real-world modelling is challenging due to the small number of registered extreme events and/or the excessive computational resources required by the hydraulic/hydrodynamic simulation models to estimate flood damage. This research has used simplified analytical functions to represent this component in each simulated case. That has made the integrated framework versatile, which allows the full range of potential flood events and warning decision situations, and the economic consequences of the resulting responses and impacts to be explored. The social aspect of how flood plain residents respond to flood warnings has been embedded in flood damage mitigation functions published in the literature.

In the simulated generic fluvial case, in Chapter 5, the forecast and observed flood impact in the at-risk community was simulated by emulating the generation of forecast and observed flood maps through a pre-assumed impact curve that estimates the percentage of floodplain properties affected by different floods. Thus, the impact was evaluated in terms of affected houses (warned and flooded houses), which was used to estimate the flood warning reliability of a FEWS based on the criterion named in this thesis as the floodplain property criterion FPC. In Chapter 6, this curve was also used to estimate the forecast and observed flood impact of an at-risk community located in a floodplain area. In this case, the flood impact was estimated in terms of flood damage and was the basis for estimating the economic effectiveness of a FEWS based on real-time flood maps. Impact curves defining the number of affected houses have been used before for doing a flood risk assessment at a national scale (Sayers *et al.*, 2015, 2018); however, to the best of the author's knowledge, they have not been used for analysing FEWSs.

In the simulated flood-prone polder system case study, an impact curve was also used to estimate the inundated area in the polder as a function of the resulting waterlogging depth (W). An impact curve can be built with relatively few flood events, and this thesis has shown that this curve type could be used to emulate rationally the process done by inundation models and contributes to the integrated framework's versatility in evaluating the benefits of a FEWS.

The response to flood warnings in the simulated generic fluvial case was simulated by using the socio-economic response functions proposed by Carsell *et al.* (2004) for houses in California, which represent the damage to the residential content after movable assets have been raised and/or evacuated. These functions assume that the residents respond to the flood warnings in a rational manner by seeking to minimise damage to their property. In reality, flood warnings may not reach the residents; they may not respond as they should, etc. These aspects have not been represented in the simulations.

The response to flood warnings in the flood-prone polder system case study was, on the other hand, simulated through pumping decision rules that represented proactive and reactive actions controlling the pumping scheme operating in the polder system. Since polder systems have not been studied before in a flood warning context, this research proposed pumping rules representing pumping strategies under different forecast scenarios. These pumping strategies and the resulting waterlogging W in the polder area after they have been done were simulated based on the simple stylized polder runoff simulation model proposed by Gao *et al.* (2008). This model has not been applied before in a flood warning context and was linked with an impact curve to represent the response and impact component (RIC) in the integrated framework.

The use of impact curves and analytical expressions or functions representing the RIC's main processes made the integrated framework versatile. Thanks to this versatility, the long-term performance of a FEWS under stationary conditions could be explored through a MC simulation approach, which would be impossible within a real-world modelling framework.

8.1.5 Factors influencing the performance of a FEWS

The MC simulation experiments conducted in this thesis identified several important factors influencing the performance of a FEWS. This was done by doing a sensitivity analysis (SA) of several parameters that define the model representing the FEWS in each simulated case. Several findings are in line with some relevant literature, whereas others represent a new contribution of this research to the overall understanding of the performance of a FEWS in each simulated case.

8.1.5.1 Simulated generic fluvial case

Reliability of flood warnings

Flood warning reliability was one of the performance attributes explored in the simulated generic fluvial case, the analysis of which considered a FEWS with and without real-time flood maps. Flood warning's reliability was defined by the metrics known as the probability of detection (POD) and false alarms ratio (FAR) evaluated based on the flooding threshold-based criterion FTC, for the FEWSs without real-time flood maps, or floodplain property-based criterion FPC, for the FEWSs *w*ith real-time flood maps. The FTC is a well-known criterion used in several research works (Verkade and Werner, 2011; Pappenberger *et al.*, 2015; Bischiniotis *et al.*, 2019), whereas the FPC is a contribution of this research which gives a more consistent estimation of flood warning reliability because it is estimated based on warned and flooded houses. The results show that the performance based on the FPC is lower than that obtained based on the FTC. That difference is directly related to the uncertainty of the flood magnitude, i.e., the difference between y and $\hat{y}$ which define the difference between the warned and flooded properties. Note that a hit in the FTC does not mean a hit in the FPC; due to the difference between these two variables, a hit in the former might include misses or false alarms in the latter (Figure 5.4a). These results suggested that i) the uncertainty of the flood magnitude is an important factor influencing the reliability of flood warnings, and ii) if inundation level forecasting is undertaken (FPC), improved forecasts would be needed to achieve the same performance level as for FTC.

The warning strategy is also one of the important factors influencing the reliability of flood warnings. This research has shown that a deterministic-based warning strategy in the FEWS produces sub-optimal decisions and that a probabilistic-based warning strategy, where the forecast errors are acknowledged, can use an optimization criterion to improve the reliability of flood warnings. This is a well-known concept and has been demonstrated in several research works (Verkade and Werner, 2011; Pagano *et al.*, 2014; Economou *et al.*, 2016). The results showed that an optimal warning strategy based on the maximum POD minus FAR optimization criterion tends to deliver higher POD and FAR values than those based on deterministic forecasts. Thus, this strategy can be used when one wants to give more

weight to the reduction of missed events whose economic consequences, in terms of floods, far overcome those associated with false alarms. FAR's target values are usually between 0.2 and 0.5, and it is the verification score of interest to humanitarians (Jolliffe and Stephenson, 2012; Bischiniotis *et al.*, 2019).

The forecasting lead time τ was also identified as an important factor influencing the flood warning reliability. As expected, flood warning reliability declines with lead time according to the performance function used, which shows a faster decline in correlation for lead times greater than the catchment lag, reflecting the greater uncertainty resulting from quantitative precipitation forecast (QPF). In particular, probabilistic forecasting, based on the maximum POD minus FAR optimization criterion, copes much better with the increasing uncertainty than deterministic forecasting, where the POD values are much higher for the probabilistic case but at the expense of higher FAR values.

The biases in the forecast mean and variance were also identified as important factors influencing the flood warning reliability. Remarkable robustness to biases in these two variates has been observed for probabilistic forecasting that is a consequence of using PU, which is based on the conditional density of y given $\hat{y}$, whereas deterministic forecasting, which is just based on $\hat{y}$, shows high sensitivity. The results for the latter case suggest that increasing the variance of the forecasts relative to those of the observed (models fitted by least squares lose variance) could improve reliability in this case.

Some research works indicate that an improvement of 10% per decade in the forecast performance is achievable in FEWSs (Pappenberger *et al.*, 2015). Through a sensitivity experiment (section 5.7.2.2), this thesis has shown that if the forecast performance is mainly controlled by $\rho_{y\hat{y}}$ (a bias in the mean or variance of the forecasts can also affect performance), these achievable benefits, in terms of flood warning's reliability, are more noticeable in FEWSs with relatively small than considerable forecast uncertainty. In line with this, some indicative results for the case of a hypothetical Morpeth FEWS suggest that a correlation of 0.94 in peak water levels would be needed to obtain POD and FAR values in the levels defined in this thesis as Medium to High category (section 5.7.2.1) for probabilistic forecasting.

Economic effectiveness

Factors influencing the economic effectiveness (E_w) of a simulated (deterministic or probabilistic) FEWS based on real-time flood maps were also explored. The economic effectiveness E_w was defined based on the expected benefits of a FEWS relative to the no warning scenario. A hydro-economic expected damage (ED) model (Figure 6.1a) was constructed that generates ED values for (a) a perfect forecast warning and (a) no warning scenario to define the reference points for measuring E_w for a FEWS based on imperfect forecasts. Thanks to the integrated framework's versatility, the economic consequences of PU in all possible situations could be included in the analysis. These economic consequences included the cost and unmitigated damage of a "wrong" flood warning decision *and* the "unnecessary" cost of the warning response or unmitigated damage after a "good" decision, i.e., when a flood is preceded by a warning (Figure 6.2a and Figure 6.2b). This analysis represents a contribution of this research since other research works in this field (Verkade and Werner, 2011; Bischiniotis *et al.*, 2019) have considered only the economic consequences of "wrong" flood warning decisions and neglected the economic consequences of flood magnitude uncertainty which is produced due to the difference between warned and flooded houses.

The results have shown that, like flood warning reliability, the warning strategy is an important factor influencing E_w of a (deterministic and probabilistic) FEWS. An optimal warning strategy based on probabilistic forecasts and an estimation of the economic benefits produced by the FEWS for each potential flood could increase the deterministic forecast-based E_w. Furthermore, it was observed that the cost of the warning response C_w controls flood warning reliability and the warning strategy in a probabilistic FEWS where warning decisions are based on optimising a probabilistic threshold. Depending on the value of C_w, an optimal warning strategy of the system could be warning more or less frequently. If C_w is low, an optimal economic strategy could suggest having high FAR values. Finally, it was shown that the benefits of using a probabilistic warning strategy with respect to a deterministic one are most noticeable in FEWSs based on forecasts with relatively high PU and where the potential economic benefits of the system are relatively high (section 6.7.2.3).

The forecasting lead time τ was also found to be an important factor influencing the E_w of a (deterministic and probabilistic) FEWS. Based on the hydro-economic ED model, and assuming that the forecast errors and uncertainty are only controlled by $\rho_{y\hat{y}}$, this research explored how E_w of a (deterministic and probabilistic) FEWS with a relatively small value of C_w as should be the case in reality, changes as τ increases. The results showed that the E_w of a (deterministic and probabilistic) FEWS reduces from the maximum E_w one can obtain from the system, i.e., that obtained from a perfect forecast scenario, as τ changes. That occurs because the PU and its economic consequences increase as the τ increases, and, therefore, its (negative) impact on E_w also increases. This E_w behaviour is in line with results obtained by Verkade and Werner (2011). However, since they assume that C_w is independent of the magnitude of the forecast flood and only consider the economic consequences of "wrong" flood warning decisions, their results overestimate E_w.

Furthermore, based on the particular analysis mentioned above (low value of C_w and forecast errors controlled only by $\rho_{y\hat{y}}$), it was found that the E_w of a (deterministic and probabilistic) FEWS does not necessarily increase when increasing τ, even though the mitigation time increases. When these FEWSs reaches considerable PU, the (negative) impact of this predictive uncertainty's economic consequences made E_w decreases as the damage is increasing. Thus, in an imperfect FEWS, it was observed that there is an optimal forecasting lead time τ that maximises the E_w of the FEWS; this lead time represents the balance between an adequate time to act and a reasonably good forecast. That E_w behaviour was found by Bischiniotis *et al.* (2019) for a probabilistic FEWS; however, there is a difference between that research work and the analysis done here; Bischiniotis *et al.* (2019) use a fixed warning strategy for the probabilistic FEWSs, whereas in this research, thanks to the versatility of the integrated framework, the warning strategy of the probabilistic FEWSs was obtained for each forecasting lead time τ, which looks for the best strategy according to τ and its associated PU. For example, in this particular analysis (low value of C_w and forecast errors controlled only by $\rho_{y\hat{y}}$), it was observed that, due to the low value of C_w, when the PU was considerable (τ > 6 hours), the best economic strategy was to increase the level of warning, accepting or "sacrificing" the low economic impact of warning more houses than one should warn, to avoid the high economic impact of having flooded houses not being warned. The results of this research, thus, lead to the conclusion that, in a FEWS,

a probabilistic-forecast-based-optimal warning strategy should be set for each lead time τ since PU, and its associated economic consequence in terms of net damage, increase as τ increases.

It is worth mentioning that, in the particular analysis mentioned above, the optimal warning strategies of the probabilistic FEWSs produced high values of FAR for FEWS with considerable PU (τ > 6 hours), which might generate the loss of credibility of the flood warning system ('cry wolf' effect). The 'cry wolf' effect might decrease the effectiveness of the proactive action, which is represented in the hydro-economic ED model by the parameter α. The value of this parameter was fixed for each lead time τ and is independent of the FAR value. Thus, the E_w of the probabilistic FEWS associated with considerable uncertainty (Figure 6.8) might be, thus, overestimated. However, Barnes *et al.* (2007) advocate that there is little evidence that a high value of FAR causes users to ignore warnings of severe flood events.

It was found that the E_w of a FEWS, based on residential contents moved and evacuated, a 6-hour lead time, and low cost of the warning response (10% of the forecast E_w), can reach approximately 15% in a perfect warning scenario (perfect forecast and response). This value cannot be achieved by imperfect FEWS due to the economic consequences of PU and the inefficiency of the proactive action (represented by the parameter α in the hydro-economic ED model). Furthermore, it was analysed that if one considers an efficiency of 70%, the E_w of an imperfect FEWS can reach 5.5 and 7.5 % for the deterministic and probabilistic forecasts, respectively (section 6.7.2.3). Furthermore, it was found that a FEWS could have good flood warning reliability but low E_w due to the poor performance of the proactive action. In this sense, it was concluded that the performance of the proactive action is an important factor influencing the E_w of a FEWS.

The values of E_w mentioned above approach that, for example, suggested by Priest *et al.* (2011) (5%, see Table 2.3); note, however, that E_w of a FEWS based on this proactive action depends on the response performance, lead time, and warning strategy (which is associated with the forecast type used). If the efficiency of 50% for the proactive action and a low cost of the warning response C_w are considered, the E_w values of deterministic FEWSs can range from 1 to 4% for lead times

between 1 and 24h. These E_w values can be improved by probabilistic FEWS, ranging from 3% to 5.5 % (section 6.7.2.1).

8.1.5.2 Flood-prone polder system case study

For the Nanjing flood-prone polder system case study, the polder manager pumps water from the inner rivers to the outer rivers to enable water to drain from the polder areas into the inner rivers. In such a flood-prone polder system, the pumping capacity is usually lower than the drainage capacity (maximum runoff entering the inner rivers). If these polder systems are operated based on reactive pumping actions (water pumped based on the observed runoff (inflow) entering the inner rivers), the probability of the inner rivers' storage capacity being overwhelmed increases. When this occurs, the runoff cannot drain freely to the inner rivers, and the waterlogging, and, therefore, the flood impact on the polder area, increases. The duration of waterlogging and the maximum inundated area are, for example, variates that measure damage which increases when this critical condition occurs in the polder system. Since the polder system's storage capacity is defined by the water level of the inner rivers when the storm arrives, flood warnings can provide time in advance to decrease that water level (increase the storage capacity) and avoid this critical condition situation. Based on this concept, a FEWS can be designed and implemented, which is characterised by a forecasting model providing, for example, forecasts of future storms, warning decisions made based on those forecasts, and responses defined by the pumping scheme operating in the flood-prone polder system. FEWSs operating for flood-prone polder systems of this kind have not, to the author's knowledge, been studied and designed before. Urban polders, particularly in the region of the case study, have been studied in another context, for example, by analysing how they impact the hydrology of the adjacent outer rivers during extreme events (Gao *et al.*, 2017; Fang *et al.*, 2018) or how they impact the flood risk of areas located downstream of these rivers (Gao *et al.*, 2018). In this sense, the FEWS proposed to mitigate the waterlogging risk in the Shazhou polder located in Nanjing, China, is a significant contribution in this field and is timely, as more intense rainfall is being experienced in recent years. The averages of the waterlogging duration ($\bar{d}_w$) and maximum inundated area ($\overline{MIA}$) in the polder were used as performance measures for this system, and the following factors influencing these measures were identified.

Since flood warnings drive the pumping actions, the warning and pumping strategies adopted in the FEWS were important factors controlling the performance of the system. Warning decisions were made based on a rainfall threshold RT curve, and the pumping strategy considered proactive and reactive response actions. The former action is triggered by the flood warning and defines the volume of water pumped before the storm arrives, whereas the reactive action defines the volume of water pumped during the storm. The volume of water pumped in the proactive action is based on an estimation of the excess runoff, defined as a proportion α of the forecast of the total runoff generated by the forecasted 'critical storm', defined as a storm that will bring the level of the inner rivers to a critical condition. Since this total runoff varies according to the characteristics of the forecasted critical storm, the performance of the FEWS was explored by using several pre-assumed values of α in the MC simulation of the system. Furthermore, since α impacts on the pumping costs, the average pumping costs ($\bar{C}_p$) were also included in the analysis. Results showed a very nice trade-off between $\bar{C}_p$ and $\overline{MIA}$ or $\bar{d}_w$, which defined a Pareto curve. Depending on the polder manager's attitude to risk, he/she can choose a value of α that represents an acceptable tradeoff between waterlogging and pumping cost.

In addition, it was observed that the performance measures that assume warning and pumping strategies based on deterministic forecasts can be improved through the use of probabilistic forecasts. Deterministic forecast-based results were improved by considering warning and pumping strategies based on probabilistic forecasts of the rainfall and runoff volume, respectively, of the predetermined critical storm.

8.1.6 The main contributions of the thesis

In the prior subsections, some important methodological contributions of this research were identified that are generic and can be implemented on any flood-prone fluvial or polder system. This section gives a summary of them so that the reader can confirm his/her understanding of the relevance of this research. These contributions are detailed below.

- ✓ A versatile generic framework based on MC simulation has been designed that can be used to assess the performance of an end-to-end FEWS in terms of reliability and economic efficiency without restrictive assumptions and

to evaluate the sensitivity of these performance measures to a range of controlling factors.

- ✓ The novel architecture of the MCFG facilitates the generation of large bivariate samples of peak water levels and their forecasts using different distributions. The predictive uncertainty PU is controlled by a correlation/lead-time performance function that represents the increase in PU with a lead time for the simulated generic fluvial case.

- ✓ A novel probabilistic forecast-based warning criterion that exploits real-time flood inundation maps has been proposed and evaluated. The floodplain property criterion FPC has been used to represent warned and flooded houses based on an impact curve that is a surrogate for flood inundation maps. This allows warnings to be issued only to properties that are forecast to be flooded, rather than all properties, and POD and FAR to be evaluated on this basis.

- ✓ A novel methodology to include the economic consequences of flood magnitude's uncertainty in the economic analysis of a FEWS for a generic fluvial case was described. It was also observed that the flood magnitude's uncertainty is an important factor to be considered when estimating the performance of a FEWS.

- ✓ The principles for the design of a FEWS operating for a flood-prone polder system were stated.

- ✓ A novel methodology for building a rainfall threshold (RT) curve for flood warning purposes in a flood-prone polder system case study was designed.

- ✓ In the simulated flood-prone-polder system case study, the results were derived to show a trade-off between the average pumping costs $\bar{C}_p$ and the measures defining the performance of the FEWS ($\overline{MIA}$ and $\bar{d}_w$).

The following conclusions cover the understanding and contributions to knowledge gained from applying the generic framework and its novel elements summarized above to a generic fluvial case and the flood-prone polder case study. The above innovations are not repeated below.

8.2 Conclusions

This thesis has opened up a new area of flood warning science that has parallels with seminal work performed by Hosking and Wallis (1988, 1997). They demonstrated how MC sampling can be used to explore the sensitivity of the estimate of T-year flood to various factors and assumptions. For example, MC simulation was used to assess the simplifying assumption that flood records from different sites are statistically independent when performing a regional flood

frequency analysis. By stochastically generating multisite annual maximum flood series, with correlations between sites, it was demonstrated that, even in the presence of intersite dependence, the bias in flood quantile estimates is unchanged. Controlled experiments employing MC simulation allow the exploration of the entire space of the problem domain, from which robust conclusions can be drawn.

The main conclusions of this research are emphasised below, according to the research objectives.

Objective 1: To design a flexible generic simulation framework that can represent the forecast, warning and response/impact components of a FEWS.

A flexible generic framework has been designed that incorporates the innovative elements summarised in section 8.1.6 above. This has enabled the overall aim of the thesis to be achieved, which is: **to build a generic framework that can simulate and identify important factors controlling the performance of a FEWS monitoring and warning a flood-prone area.**

Objective 2: To design a Monte Carlo flood and forecast generator (MCFG) applicable to a generic fluvial case and a flood-prone polder system case.

- ✓ The architecture used to build the MCFG allows PU to be controlled by a lead-time-correlation function that links a forecasting lead time with PU in the simulated generic fluvial case. Several bivariate distributions and MC generation algorithms can be employed to generate large samples of forecast and observed water levels for the generic fluvial case.

- ✓ Based on the univariate analysis of peak water levels for four gauging stations in the North of England, the five-parameter BED, the seven-parameters BLND, and the seven-parameter BGM were considered to be suitable models to represent these data.

- ✓ A bivariate analysis of observed peak water levels and their forecasts derived from a previously calibrated model for one of the four stations showed that the same distribution could be used to describe both observed and forecast levels. This allowed any of the three distributions identified above to be employed for bivariate simulation.

- ✓ For the Nanjing polder case study, a univariate analysis of observed and forecast daily rainfalls showed that the three-parameter Gamma distribution fitted the data well, suggesting, thus, that the seven-parameter BGM was a suitable model to represent the bivariate relationship of these variates.

- ✓ These results have suggested that, when using a parametric bivariate model to represent the predictand and predictor based on the approach described in this research, one should try to use a model that incorporates an analytical expression that links its dependence structure in the transformed domain with that in the real domain. By doing that, the MCFG and the integrated framework gain versatility.

Objective 3: To design the flood warning decision component (FWDC) of a FEWS, and to simulate and identify important factors controlling the flood warning reliability of a FEWS for a generic fluvial river case under deterministic and probabilistic forecast information.

- ✓ Predictive uncertainty PU, estimated from the predictive density of the observed value conditional on the forecast, is an essential factor to be considered when estimating the flood warning reliability of a FEWS. Probabilistic forecasting that employed this concept outperformed deterministic forecasting across all the sensitivity experiments conducted in this thesis. This was achieved by optimizing a probabilistic threshold to maximise flood warning reliability, characterised by the metrics POD and FAR.

- ✓ The warning strategy is also one of the important factors influencing the reliability of flood warnings. This research showed that a deterministic-based warning strategy in FEWS operation produces sub-optimal decisions and that a warning strategy based on optimising a probabilistic threshold, where the forecast errors are acknowledged, can improve the reliability of flood warnings.

- ✓ In this research, the maximum difference between POD and FAR was used as an optimization criterion for probabilistic forecasting. The results showed that an optimal warning strategy based on this criterion delivers POD and FAR values greater than those based on deterministic forecasts. Thus, this strategy can be used when one wants to give more weight to the reduction of missed events whose economic consequences, in terms of floods, far overcomes those associated with false alarms.

- ✓ The use of the floodplain property-based criterion FPC results in lower values of POD and FAR than the traditional flooding threshold-based criterion FTC, and is, therefore, a more demanding reliability criterion. Impact curves defining the number of affected houses can be used for exploring the flood warning's reliability in terms of affected houses. An impact curve can be built with relatively few flood events, and it can exploit very well the information produced by inundation models and contributes to the integrated framework's versatility

- ✓ The correlation between the observed and forecast peak water levels has been shown to be the most important factor controlling flood warning reliability, which improves as the correlation increases. The simulation framework used here allows the level of correlation needed to deliver a

target reliability to be estimated a priori, before any investment is made in improving an existing model or developing a new model.

- ✓ As the lead time increases beyond the catchment lag, and the uncertainty grows due to the use of QPFs, the decline in correlation results in decline in POD and FAR to poor levels for deterministic forecasting. However, with probabilistic forecasting, used to describe this POD and FAR both increase using the maximum POD minus FAR optimization criterion, reflecting a strategy that avoids misses at the expense of more false alarms as noted above.

- ✓ Negative biases in the mean and variance, i.e., underestimation, lead to a decline in the reliability of deterministic flood warnings, whereas the reliability of probabilistic forecasts based on optimising a probabilistic threshold is robust to such biases. This is a remarkable property of probabilistic forecasts that have not been clearly demonstrated before.

Objective 4: To design the response and impact component in the FEWS for each simulated case, and to simulate and identify important factors controlling the economic effectiveness of a FEWS for a generic fluvial river case under deterministic and probabilistic forecast information.

- ✓ Impact curves and analytical expressions or functions representative of US socio-economic conditions were used to represent the RIC's main flood damage processes for the generic fluvial case, which has made the integrated framework versatile, as these can be modified to represent socio-economic conditions in other countries. Thanks to this versatility, the long-term socio-economic performance of a FEWS can be explored through a MC simulation approach, which would be impossible within a real-world modelling framework.

- ✓ The RIC component for the generic fluvial case exploits the novel use of the FPC criterion introduced in this thesis. By combining a house-based flood impact curve that is a surrogate for flood inundation maps with a house-based family of damage functions, the long-term economic effectiveness of a FEWS can be evaluated.

- ✓ Thanks to the integrated framework's versatility, the economic consequences of utilising PU in decision-making could be included in all possible situations in the analysis. Other research works in this field have considered only the economic consequences of "wrong" flood warning decisions and neglected the economic consequences of flood magnitude uncertainty which is produced due to the difference between warned and flooded houses.

- ✓ Although damage mitigation from proactive action should increase with lead time and mitigation time, the E_w of (deterministic and probabilistic) FEWSs does not necessarily increase when increasing τ. When imperfect FEWSs reached considerable PU, the (negative) impact of this predictive uncertainty's economic consequences make E_w decreases. Thus, in an

imperfect FEWS, it was observed that there is an optimal forecasting lead time τ that maximises E_w; this lead time represents the balance between an adequate time to act and a reasonably good forecast.

- ✓ The results showed that, as for the flood warning reliability performance measure, the warning strategy is an important factor influencing E_w. Based on the particular analysis (low value of C_w(γ=0.10), forecast errors controlled only by $\rho_{y\hat{y}}$, and, an efficiency of 50% for the proactive action (α=0.5)), it was found that E_w based on deterministic-forecast-based warning decisions can range from 1 to 4% for lead times between 1 and 24h. These E_w values were improved by probabilistic warning decisions, ranging from 3% to 5.5 %.

- ✓ The results showed that the benefits of using a probabilistic warning strategy with respect to a deterministic one are most noticeable in FEWSs based on forecasts with relatively high predictive uncertainty and where the potential economic benefits of the system are relatively high.

- ✓ The cost of the warning response C_w has been shown to control the flood warning reliability and the warning strategy in a probabilistic FEWS where warning decisions are based on a probabilistic threshold. Depending on the value of C_w, an optimal warning strategy of the system could be warning less or more frequently.

- ✓ It was concluded that the performance of the proactive action is the most important factor influencing the E_w of a FEWS. It was found that a FEWS could have good flood warning reliability but low E_w due to the poor performance of the proactive action.

Objective 5: To apply the generic framework to a case study of the operation of a flood-prone polder system in Nanjing, China

- ✓ The design principles for a FEWS operating in a flood-prone polder system have been formulated. The FEWS proposed to mitigate the waterlogging risk in the Shazhou polder located in Nanjing, China, is a significant contribution in this field.

- ✓ The pumping strategies and the resulting waterlogging W in the polder area were simulated based on a simple stylized water balance model. This model has not been applied before in a flood warning context and was linked with an impact curve to represent the response and impact component (RIC) in the integrated framework.

- ✓ The warning and pumping strategies adopted in the FEWS were important factors controlling the performance of the system. Deterministic forecasts-based strategies produced performance measures that could be improved by considering warning and pumping strategies based on probabilistic forecasts of the rainfall and runoff volume, respectively, of the impending critical storm.

- ✓ The proactive action was formulated such that a parameter α representing the amount of proactive pumping could be used to define a Pareto tradeoff curve between average waterlogged area and pumping cost. It was shown that probabilistic forecasts could outperform deterministic forecasts by generating points closer to Utopia on the Pareto curve.

- ✓ The Nanjing authority responsible for the operation of the Shazou and other polders in the city could gain considerable operational benefits in managing flooding by adopting the FEWS design explored in this research.

8.3 Recommendation for future work

This is an appropriate juncture at which to consider the limitations of the generic framework and its constituent components; these can form the basis of some further work. For example, important assumptions made in the framework are that forecast performance does not change in the future, i.e., stationarity is assumed, and that warning decisions are made based on a predefined decision rule. In addition, the representation of the response and impact process through analytical expressions can neglect some physical and social processes in the forecast of the flood impact and in representing the response and impact processes. Therefore, the models and analyses described in this thesis for each simulated case can be subject to several improvements or adaptations that could be done in future work. These are discussed in turn below.

At the beginning of this thesis, the overarching issue of climate change was mentioned in the context of a FEWS having a very important role to play in managing the increasing flood risk. A 10% improvement per decade in FEWS performance has been set (Pappenberger *et al.*, 2015) as a general target to be aimed for which should contribute to managing the increasing flood risk, and the research conducted here has provided some insight into how this might be achieved. However, the impact of changes in rainfall and flood extremes on FEWS performance is a specific topic that should be investigated in further work. Guerreiro et al. (2018) have shown how the characteristics of hourly rainfall extremes are changing at the continental scale, and these changes could be embedded in RainSim to explore, for specific catchments such as the Eden for which a calibrated SHETRAN model is available, how the performance of a FEWS might change under these transient conditions, and how the estimation of PU and flood warning decision rules should adapt. Furthermore, significant changes in

hourly rainfall extremes associated with urbanisation and the heat-island effect in mega-cities are also being observed (Li et al., 2020). The parameters of a bivariate simulation model could be adapted fairly simply to reflect indirectly these changes, e.g., by changing the parameters of the marginal distribution of y while leaving those of $\hat{y}$ unchanged.

The MCFG could be extended to incorporate continuous simulation in which rainstorms and their forecasts generated by RainSim are used to derive long-term continuous simulations of flood hydrographs and their forecasts. Such simulations would provide considerable additional information for the assessment of impacts but at a high computational cost. The use of the simpler MCFG developed here with selective sampling from the continuous simulation approach could overcome this difficulty. This would also support the work suggested above.

In all the sensitivity analysis experiments carried out in this thesis, perfect knowledge of the population of flood events and their forecasts has been assumed. In a real-world setting, the bivariate distribution of peak levels and their forecasts would have to be built to estimate PU. MC sampling experiments should be conducted to establish the sample sizes needed to estimate the PU reliably and the consequent FEWS reliability and economic effectiveness.

The sensitivity analysis, SA, conducted in the simulated generic fluvial case to identify the influence of parameters (factors) on the performance of a FEWS was based on the one-at-a-time (OAT) method. This analysis, in essence, represented a local SA of the model representing the FEWS. The local SA does not allow interactions among the model's input parameters to be detected and how these interactions are related to the performance of the model. This analysis can be done based on a global SA of the model, which explores the interactions among the parameters. The use of scatter plots or coloured scatter plots are useful tools for these purposes (Pianosi *et al.*, 2016).

In the simulated generic fluvial case, the hydro-economic ED model was used to explore the economic effectiveness of a FEWS as a function of the forecasting lead time τ, and the results showed that for a FEWS with a low value of the cost of the warning response C_w, an (optimal) probabilistic-forecast-based warning strategy produced high values of FAR, which might generate a loss of credibility of the flood

warning system (‘cry wolf’ effect). The ‘cry wolf’ effect might, in turn, decrease the effectiveness of the proactive action, which is represented in the hydro-economic ED model by the parameter α that is an input parameter and is independent of the FAR value. In this context, even though the negative impact of the ‘cry wolf’ effect on the FEWS is an arguable topic (Barnes *et al.*, 2007), it would be interesting to link the parameter α with the FAR value through an internal model function to connect the proactive action's effectiveness from the ‘cry wolf’ effect. Another improvement of this model would be to include the householders' decisions, which in essence, would be to represent the decisions in the response of the FEWS, which is something considered in the simulated flood-prone polder system case study.

The performance of the FEWS designed for the flood-prone polder system case study based on different pumping strategies is something that can also be explored in future work. The pumping strategy adopted is characterised by a proactive and reactive action, where the latter action is conducted before the storm arrives and the latter action during the storm, which assumes that the water is pumped based on the observed runoff (inflow) entering the inner rivers. Another pumping strategy would be, for example, to consider that the proactive action is conducted during the storm in which the water is pumped at a different rate to the observed runoff entering the inner rivers. The pumping rate could be defined by assuming that updated real-time forecasts are provided to the polder manager from time to time. Pumping strategies based on the forecast of the rainfall profile and volume of the impending critical storm is another option to be explored.

Based on the design of the Shazou polder FEWS, a serious game could be designed that would allow the managers of polder flooding in Nanjing the effectiveness of different proactive and reactive actions to be explored virtually using the Pareto curve, as a way of enhancing confidence about the potential benefits that could be gained, and establishing an acceptable balance between waterlogging and pumping cost.

The approach used to represent probabilistic forecast-based flood warning decisions in each simulated case was based on a probabilistic threshold PT approach. Based on this approach, the probability of exceedance, obtained by integrating the predictive density, i.e., PU, above a predefined level, is compared with an optimised

PT that triggers the decision. Probabilistic warning decisions can also be represented based on a Bayesian decision scheme where the predictive density and a utility/loss function are used to set rules based on expected values (Todini, 2017). Bayesian FEWSs have been designed based on rainfall (Martina *et al.*, 2006; Economou *et al.*, 2016) or water levels (Krzysztofowicz, 1993). In this sense, exploring the performance of the simulated FEWSs based on Bayesian flood warning decisions is something that can be done in future work.

Another improvement to the proposed generic framework would be to include the human component in the simulation of the FEWSs to explore the influence of human behaviour on its performance. The human behaviour in FEWSs has been simulated based on analytical expressions (Ferrell and Krzysztofowicz, 1983; Girons Lopez *et al.*, 2017) or agent-based models (ABMs) (computational methods that simulate autonomous decision-making entities' actions and interactions in a system)(Dawson *et al.*, 2011; Du *et al.*, 2016; Yang *et al.*, 2018; O'Shea *et al.*, 2020). The framework components built in this thesis are in a very good position to obtain relatively rapid results through the first approach and to accommodate a more complex framework extension based on ABMs.

Finally, in the Shazou polder FEWS, simulated rainfall generated from RainSim was used to represent forecasts and explore what might be gained by using forecasts in reality. Thus, the assumed statistical properties defining the relationship between the observations and their forecasts are not based on real-world records, and one can argue that they represent a very positive situation. For example, the special correlation coefficient defining the predictive uncertainty PU can be considered too optimistic (0.93). In this context, a future work could be to use real-world pair forecast and observed values in this analysis to explore how the outcomes shown in this work behave under (actual) different statistical properties defining the observation-forecast relationship. This real-world data-based analysis would involve two basic steps: i) to use the hourly and daily statistics of the observed daily rainfall and their forecasts to set the parameters in RainSim, and ii) to define the joint distribution of the observed daily rainfall and their forecasts based on the real-world records.

The use of "simulation mode" forecasts in this thesis can be considered a valid approach to explore the benefits of using longer lead time forecasts in FEWSs

where the effect of updating has died out. This simulated method-based approach also responded to the present-day reality about the lack of or limited records of observed extreme values and their forecasts in operational flood forecasting systems. If real-world data were the basis of this thesis, many of the methodological contributions given in this work would not be possible to achieve because the lack of data does not allow one to study the full FEWS chain. Despite this, it is also true that flood warning/forecast validation databases are starting to be available for operational flood forecasting systems (Ayzel *et al.*, 2019; Harrigan *et al.*, 2020), and the number of registered extreme events is increasing. Therefore, the imperative next step to be done in this work would be to use one day that information in the methodologies suggested here to explore the FEWS performance. Since these methodologies are mainly based on the statistical relationship between the predictand and predictor, the information generated from any flood forecasting technique can be used, such as forecasts generated from rainfall-runoff and/or flooding routing models or by advanced data-driven models, e.g., machine learning models. However, it is the case that forecasts generated by operational agencies are not always archived or made available for analysis in external research, but the framework can still demonstrate to these agencies what target correlations are needed to achieve the required performance as measured, for example, by POD and FAR.

References

Adams TE. 2016. Flood Forecasting in the United States NOAA/National Weather Service. *Flood Forecasting*: 249–310 DOI: 10.1016/B978-0-12-801884-2.00010-4

Adamson PT, Zucchini W. 1984. On the application of a censored log-normal distribution to partial duration series of storms. *Water S.A* **10** (3): 136–146

Ahilan S, O'Sullivan JJ, Bruen M, Brauders N, Healy D. 2013. Bankfull discharge and recurrence intervals in Irish rivers. *Proceedings of the Institution of Civil Engineers: Water Management* **166** (7): 381–393 DOI: 10.1680/wama.11.00078

Alfieri L, Thielen J, Pappenberger F. 2012. Ensemble hydro-meteorological simulation for flash flood early detection in southern Switzerland. *Journal of Hydrology* DOI: 10.1016/j.jhydrol.2011.12.038

Andrews ED. 1980. Effective and bankfull discharges of streams in the Yampa River basin, Colorado and Wyoming. *Journal of Hydrology* **46** (3–4): 311–330 DOI: 10.1016/0022-1694(80)90084-0

Arheimer B, Lindström G, Olsson J. 2011. A systematic review of sensitivities in the Swedish flood-forecasting system. *Atmospheric Research* DOI: 10.1016/j.atmosres.2010.09.013

Arnal L, Anspoks L, Manson S, Neumann J, Norton T, Stephens E, Wolfenden L, Cloke HL. 2020. "Are we talking just a bit of water out of bank? Or is it Armageddon?" Front line perspectives on transitioning to probabilistic fluvial flood forecasts in England. *Geoscience Communication* DOI: 10.5194/gc-3-203-2020

Ayzel G, Varentsova N, Erina O, Sokolov D, Kurochkina L, Moreydo V. 2019. OpenForecast: The First Open-Source Operational Runoff Forecasting System in Russia. *Water (Switzerland)* **11** (8): 1546 DOI: 10.3390/w11081546

Bačová-Mitková V, Onderka M. 2010. Analysis of extreme hydrological Events on the danube using the Peak Over Threshold method. *Journal of Hydrology and*

Hydromechanics **58** (2): 88–101

Balakrishna N, Lai CD. 2009. *Continuous bivariate distributions*. Springer Science & Business Media: New York. DOI: 10.1007/b101765

Barbetta S, Coccia G, Moramarco T, Todini E. 2016. Case Study: A Real-Time Flood Forecasting System with Predictive Uncertainty Estimation for the Godavari River, India. *Water* **8** (10): 463 DOI: 10.3390/w8100463

Barnes LR, Gruntfest EC, Hayden MH, Schultz DM, Benight C. 2007. False alarms and close calls: A conceptual model of warning accuracy. *Weather and Forecasting* DOI: 10.1175/WAF1031.1

Basher R, Page J, Woo J, Davies ML, Synolakis CE, Farnsworth AF, Steacey S. 2006. Global early warning systems for natural hazards: Systematic and people-centred. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* DOI: 10.1098/rsta.2006.1819

Bezak N, Brilly M, Šraj M. 2014. Comparison between the peaks-over-threshold method and the annual maximum method for flood frequency analysis. *Hydrological Sciences Journal* DOI: 10.1080/02626667.2013.831174

Bhunya PK, Singh RD, Berndtsson R, Panda SN. 2012. Flood analysis using generalized logistic models in partial duration series. *Journal of Hydrology* DOI: 10.1016/j.jhydrol.2011.11.037

Birkinshaw SJ, Bathurst JC, Iroumé A, Palacios H. 2011. The effect of forest cover on peak flow and sediment discharge-an integrated field and modelling study in central-southern Chile. *Hydrological Processes* DOI: 10.1002/hyp.7900

Bischiniotis K, van den Hurk B, Coughlan de Perez E, Veldkamp T, Nobre GG, Aerts J. 2019. Assessing time, cost and quality trade-offs in forecast-based action for floods. *International Journal of Disaster Risk Reduction* DOI: 10.1016/j.ijdrr.2019.101252

Blöschl G. 2008. Flood warning - on the value of local information. *International Journal of River Basin Management* DOI: 10.1080/15715124.2008.9635336

Boelee L, Lumbroso DM, Samuels PG, Cloke HL. 2019. Estimation of uncertainty in flood forecasts—A comparison of methods. *Journal of Flood Risk Management* **12** (S1) DOI: 10.1111/jfr3.12516

Bogner K, Pappenberger F. 2011. Multiscale error analysis, correction, and predictive uncertainty estimation in a flood forecasting system. *Water Resources Research* DOI: 10.1029/2010WR009137

Bogner K, Pappenberger F, Cloke HL. 2012. Technical Note: The normal quantile transformation and its application in a flood forecasting system. *Hydrology and Earth System Sciences* **16** (4): 1085–1094 DOI: 10.5194/hess-16-1085-2012

BOM. 2019. Flood Warning Services. *Flood Warning Services* Available at: http://www.bom.gov.au/water/floods/floodWarningServices.shtml [Accessed 10 May 2019]

Bowman KO, Shenton LR. 2011. Gamma Distribution. In *International Encyclopedia of Statistical Science*Springer Berlin Heidelberg; 573–575. DOI: 10.1007/978-3-642-04898-2_269

Burton A, Kilsby CG, Fowler HJ, Cowpertwait PSP, O'Connell PE. 2008. RainSim: A spatial-temporal stochastic rainfall modelling system. *Environmental Modelling and Software* DOI: 10.1016/j.envsoft.2008.04.003

Carsell KM, Pingel ND, Ford DT. 2004. Quantifying the Benefit of a Flood Warning System. *Natural Hazards Review* **5** (3): 131–140 DOI: 10.1061/(ASCE)1527-6988(2004)5:3(131)

Castro JM, Jackson PL. 2007. Bankfull discharge recurrence intervals and regional hydraulic geometry relationships: Patterns in the Pacific Northwest, USA. *Journal of the American Water Resources Association* **37** (5): 1249–1262 DOI: 10.1111/j.1752-1688.2001.tb03636.x

Choulakian V, El-Jabi N, Moussi J. 1990. On the distribution of flood volume in partial duration series analysis of flood phenomena. *Stochastic Hydrology and Hydraulics* DOI: 10.1007/BF01543085

Chow V Te. 1954. The Log-Probability Law and Its Engineering Applications.

Proceedings of the American Society of Civil Engineers **80** (11): 1–25

Claps P, Laio F. 2003. Can continuous streamflow data support flood frequency analysis? An alternative to the partial duration series approach. *Water Resources Research* DOI: 10.1029/2002WR001868

Cloke HL, Pappenberger F. 2009. Ensemble flood forecasting: A review. *Journal of Hydrology* DOI: 10.1016/j.jhydrol.2009.06.005

Cluckie ID, Han D. 2000. Fluvial Flood Forecasting. *Water and Environment Journal* **14** (4): 270–276 DOI: 10.1111/j.1747-6593.2000.tb00260.x

Coccia G. 2011. Analysis and Developments of Uncertainty Processors for Real Time Flood Forecasting.Universitá di Bologna.

Coccia G, Todini E. 2011. Recent developments in predictive uncertainty assessment based on the model conditional processor approach. *Hydrology and Earth System Sciences* **15** (10): 3253–3274

Crawford N., Linsley RK. 1966. Digital simulation in hydrology, Stanford Watershed Model IV

Crutcher HL. 1975. A Note on the Possible Misuse of the Kolmogorov-Smirnov Test. *Journal of Applied Meteorology* DOI: 10.1175/1520-0450(1975)014<1600:anotpm>2.0.co;2

Cunnane C. 1979. A note on the Poisson assumption in partial duration series models. *Water Resources Research* **15** (2): 489–494 DOI: 10.1029/WR015i002p00489

D'Agostino RB, Stephens MA. 1986. *Goodness-of-fit techniques*. Dekker.

Dale M, Wicks J, Mylne K, Pappenberger F, Laeger S, Taylor S. 2014. Probabilistic flood forecasting and decision-making: An innovative risk-based approach. *Natural Hazards* DOI: 10.1007/s11069-012-0483-z

Dawson RJ, Peppe R, Wang M. 2011. An agent-based model for risk-based flood incident management. *Natural Hazards* **59** (1): 167–189 DOI: 10.1007/s11069-011-9745-4

Day HJ. 1970. *Flood warning benefit evaluation-Susquehanna River Basin (urban residences)* (Silver Spring, ed.).

Devak M, Dhanya CT. 2017. Sensitivity analysis of hydrological models: review and way forward. *Journal of Water and Climate Change*: jwc2017149 DOI: 10.2166/wcc.2017.149

Downton F. 1970. Bivariate Exponential Distributions in Reliability Theory. *Journal of the Royal Statistical Society: Series B (Methodological)* DOI: 10.1111/j.2517-6161.1970.tb00852.x

Du E, Rivera S, Cai X, Myers L, Ernest A, Minsker B. 2016. Impacts of Human Behavioral Heterogeneity on the Benefits of Probabilistic Flood Warnings: An Agent-Based Modeling Framework. *JAWRA Journal of the American Water Resources Association*: n/a-n/a DOI: 10.1111/1752-1688.12475

Dugar S, Smith P, Parajuli B, Khanal S, Brown S, Gautam D, Bhandari D, Gurung G, Shakya P, Kharbuja R, et al. 2017. Enhancing Community Based Early Warning Systems in Nepal with Flood Forecasting Using Local and Global Models. *19th EGU General Assembly, EGU2017, proceedings from the conference held 23-28 April, 2017 in Vienna, Austria., p.8995* **19**: 8995 Available at: http://adsabs.harvard.edu/abs/2017EGUGA..19.8995D [Accessed 10 May 2019]

Economou T, Stephenson DB, Rougier JC, Neal RA, Mylne KR. 2016. On the use of Bayesian decision theory for issuing natural hazard warnings. In *Proc. R. Soc. A*The Royal Society; 20160295.

Edwards PJ, Watson EA, Wood F. 2019. Toward a Better Understanding of Recurrence Intervals, Bankfull, and Their Importance. *Journal of Contemporary Water Research & Education* **166** (1): 35–45 DOI: 10.1111/j.1936-704x.2019.03300.x

Environment Agency. 2019. Flood warning service Available at: https://flood-warning-information.service.gov.uk/station/5127 [Accessed 24 September 2019]

Environment Agency NFWC. 2003. Flood Warning Investment Strategy Appraisal

Report 2003/04 to 2012/13. Frimley-UK.

European Parliament. 2007. Directive 2007/60/EC of the European Parliament and of the Council of 23 October 2007 on the assessment and management of flood risks

Ewen J, Parkin G, O'Connell PE. 2000. SHETRAN: Distributed River Basin Flow and Transport Modeling System. *Journal of Hydrologic Engineering* DOI: 10.1061/(asce)1084-0699(2000)5:3(250)

Fang G, Yuan Y, Gao Y, Huang X, Guo Y. 2018. Assessing the effects of urbanization on flood events with urban agglomeration polders type of flood control pattern using the HEC-HMS model in the Qinhuai River Basin, China. *Water (Switzerland)* DOI: 10.3390/w10081003

Fernández-Nóvoa D, García-Feal O, González-Cao J, de Gonzalo C, Rodríguez-Suárez JA, del Portal CR, Gómez-Gesteira M. 2020. MIDAS: A New Integrated Flood Early Warning System for the Miño River. *Water (Switzerland)* **12** (9): 2319 DOI: 10.3390/W12092319

Ferrell WR, Krzysztofowicz R. 1983. A model of human response to flood warnings for system evaluation. *Water Resources Research* DOI: 10.1029/WR019i006p01467

Flack D, Skinner C, Hawkness-Smith L, O'Donnell G, Thompson R, Waller J, Chen A, Moloney J, Largeron C, Xia X, et al. 2019. Recommendations for Improving Integration in National End-to-End Flood Forecasting Systems: An Overview of the FFIR (Flooding From Intense Rainfall) Programme. *Water* **11** (4): 725 DOI: 10.3390/w11040725

Gao C, Liu J, Cui H, Doddi Y. 2008. An applicable method to calculate drainage modulus in urbanized lowlying area. In *2008 International Workshop on Education Technology and Training and 2008 International Workshop on Geoscience and Remote Sensing, ETT and GRS 2008*. DOI: 10.1109/ETTandGRS.2008.216

Gao C, Liu J, Cui H, Hu J. 2013. Treatment of pump drainage boundary in riverside city. *Environmental Earth Sciences* **68** (5): 1435–1442 DOI: 10.1007/s12665-

012-1841-6

Gao C, Liu J, Cui H, Zhu J. 2009. An effective way to determine maximum capacity of pump stations for urbanized polders. In *3rd International Conference on Bioinformatics and Biomedical Engineering, ICBBE 2009*. DOI: 10.1109/ICBBE.2009.5162314

Gao Y, Wang D, Zhang Z, Ma Z, Guo Z, Ye L. 2018. Analysis of Flood Risk of Urban Agglomeration Polders Using Multivariate Copula. *Water* DOI: 10.3390/w10101470

Gao Y, Yuan Y, Wang H, Schmidt AR, Wang K, Ye L. 2017. Examining the effects of urban agglomeration polders on flood events in Qinhuai River basin, China with HEC-HMS model. *Water Science and Technology* DOI: 10.2166/wst.2017.023

Georgakakos KP. 2006. Analytical results for operational flash flood guidance. *Journal of Hydrology* DOI: 10.1016/j.jhydrol.2005.05.009

George L, Modarres M. 1994. What Every Engineer Should Know about Reliability and Risk Analysis. *Technometrics* DOI: 10.2307/1270248

Girons Lopez M, Di Baldassarre G, Seibert J. 2017. Impact of social preparedness on flood early warning systems. *Water Resources Research* DOI: 10.1002/2016WR019387

Golian S, Fallahi MR, Behbahani SM, Sharifi S, Sharma A. 2015. Real-Time Updating of Rainfall Threshold Curves for Flood Forecasting. *Journal of Hydrologic Engineering* **20** (4): 04014059 DOI: 10.1061/(ASCE)HE.1943-5584.0001049

Golian S, Saghafian B, Maknoon R. 2010. Derivation of Probabilistic Thresholds of Spatially Distributed Rainfall for Flood Forecasting. *Water Resources Management* DOI: 10.1007/s11269-010-9619-7

Gouldby B, Samuels P. 2009. Language of Risk: Project Definitions (Second edition) Available at: http://www.floodsite.net/html/partner_area/project_docs/FLOODsite_Language_of_Risk_v4_0_P1.pdf

Guerreiro SB, Fowler HJ, Barbero R, Westra S, Lenderink G, Blenkinsop S, Lewis E, Li XF. 2018. Detection of continental-scale intensification of hourly rainfall extremes. *Nature Climate Change* DOI: 10.1038/s41558-018-0245-3

Hapuarachchi HAP, Wang QJ, Pagano TC. 2011. A review of advances in flash flood forecasting. *Hydrological Processes* DOI: 10.1002/hyp.8040

Harrigan S, Zsoter E, Alfieri L, Prudhomme C, Salamon P, Wetterhall F, Barnard C, Cloke H, Pappenberger F. 2020. GloFAS-ERA5 operational global river discharge reanalysis 1979-present. *Earth System Science Data* **12** (3): 2043–2060 DOI: 10.5194/essd-12-2043-2020

Hashino T, Bradley AA, Schwartz SS. 2007. Evaluation of bias-correction methods for ensemble streamflow volume forecasts. *Hydrology and Earth System Sciences* DOI: 10.5194/hess-11-939-2007

Heuvelink D, Berenguer M, Brauer CC, Uijlenhoet R. 2020. Hydrological application of radar rainfall nowcasting in the Netherlands. *Environment International* DOI: 10.1016/j.envint.2019.105431

Hosking JRM, Wallis JR. 1988. The effect of intersite dependence on regional flood frequency analysis. *Water Resources Research* DOI: 10.1029/WR024i004p00588

Hosking JRM, Wallis JR. 1997. *Regional Frequency Analysis*. Cambridge University Press. DOI: 10.1017/cbo9780511529443

Ibbit R, Woods R. 2003. Enhancing flood forecasts. *TEPHRA* **20**: 5 Available at: https://www.civildefence.govt.nz/assets/Uploads/publications/tephra-june-2003.pdf

Jain SK, Mani P, Jain SK, Prakash P, Singh VP, Tullos D, Kumar S, Agarwal SP, Dimri AP. 2018. A Brief review of flood forecasting techniques and their applications. *International Journal of River Basin Management* **16** (3): 329–344 DOI: 10.1080/15715124.2017.1411920

Janes V, Holman I, Birkinshaw S, O'Donnell G, Kilsby C. 2018. Improving bank erosion modelling at catchment scale by incorporating temporal and spatial variability. *Earth Surface Processes and Landforms* DOI: 10.1002/esp.4149

Javelle P, Organde D, Demargne J, Saint-Martin C, de Saint-Aubin C, Garandeau L, Janet B. 2016. Setting up a French national flash flood warning system for ungauged catchments based on the AIGA method (M Lang, F Klijn, and P Samuels, eds). *E3S Web of Conferences* **7**: 18010 DOI: 10.1051/e3sconf/20160718010

Jolliffe I, Stephenson DB. 2012. *Forecast Verification. A Practitioner's Guide in Atmospheric Science*. John Wiley & Sons, Ltd.: Exeter, UK.

Kambalimath S, Deka PC. 2020. A basic review of fuzzy logic applications in hydrology and water resources. *Applied Water Science* **10** (8): 191 DOI: 10.1007/s13201-020-01276-2

Karim F, Hasan M, Marvanek S. 2017. Evaluating annual maximum and partial duration series for estimating frequency of small magnitude floods. *Water (Switzerland)* DOI: 10.3390/w9070481

Kelly KS, Krzysztofowicz R. 1997. A bivariate meta-Gaussian density for use in hydrology. *Stochastic Hydrology and Hydraulics* DOI: 10.1007/BF02428423

Klein B, Meissner D, Kobialka HU, Reggiani P. 2016. Predictive uncertainty estimation of hydrological multi-model ensembles using pair-copula construction. *Water (Switzerland)* DOI: 10.3390/w8040125

Koo BK, O'Connell PE. 2006. An integrated modelling and multicriteria analysis approach to managing nitrate diffuse pollution: 2. A case study for a chalk catchment in England. *Science of the Total Environment* DOI: 10.1016/j.scitotenv.2005.05.013

Kottegoda NT, Rosso R. 2008. *Applied Statistics for Civil and Environmental Engineers*.

Krzysztofowicz R. 1993. A theory of flood warning systems. *Water Resources Research* **29** (12): 3981–3994 DOI: 10.1029/93WR00961

Krzysztofowicz R. 1999a. Bayesian Forecasting via Deterministic Model. *Risk Analysis* **19** (4): 739–749 DOI: 10.1111/j.1539-6924.1999.tb00443.x

Krzysztofowicz R. 1999b. Bayesian theory of probabilistic forecasting via

deterministic hydrologic model. *Water Resources Research* **35** (9): 2739–2750 DOI: 10.1029/1999WR900099

Krzysztofowicz R, Davis DR. 1983. A methodology for evaluation of flood forecast-response systems: 1. Analyses and concepts. *Water Resources Research* DOI: 10.1029/WR019i006p01423

Krzysztofowicz R, Kelly KS. 2000. Hydrologic uncertainty processor for probabilistic river stage forecasting. *Water Resources Research* **36** (11): 3265–3277 DOI: 10.1029/2000WR900108

Krzysztofowicz R, Kelly KS, Long D. 1994. Reliability of flood warning systems. *Journal of Water Resources Planning and Management* DOI: 10.1061/(ASCE)0733-9496(1994)120:6(906)

Lai CD, Rayner JCW, Hutchinson TP. 1999. Robustness of the sample correlation - the bivariate lognormal case. *Journal of Applied Mathematics and Decision Sciences* **3** (1): 7–19 DOI: 10.1155/S1173912699000012

Lane R, Coxon G, E Freer J, Wagener T, J Johnes P, P Bloomfield J, Greene S, J A Macleod C, M Reaney S. 2019. Benchmarking the predictive capability of hydrological models for river flow and flood peak predictions across over 1000 catchments in Great Britain. *Hydrology and Earth System Sciences* DOI: 10.5194/hess-23-4011-2019

Laoupi A, Tsakiris G. 2007. Assessing Vulnerability in Cultural Landscapes. In *EWRA Symposium 'Water Resources Management: New Approaches and Technologies'*Chania, Crete, Greece.

Leedal D, Weerts AH, Smith PJ, Beven KJ. 2013. Application of data-based mechanistic modelling for flood forecasting at multiple locations in the Eden catchment in the National Flood Forecasting System (England and Wales). *Hydrology and Earth System Sciences* DOI: 10.5194/hess-17-177-2013

Lewis E, Birkinshaw S, Kilsby C, Fowler HJ. 2018. Development of a system for automated setup of a physically-based, spatially-distributed hydrological model for catchments in Great Britain. *Environmental Modelling and Software* DOI: 10.1016/j.envsoft.2018.07.006

Lewis PWA, Orav EJ. 2018. *Simulation Methodology for Statisticians, Operations Analysts, and Engineers (1988)*. CRC Press: New York.

Li W, Duan Q, Miao C, Ye A, Gong W, Di Z. 2017. A review on statistical postprocessing methods for hydrometeorological ensemble forecasting. *Wiley Interdisciplinary Reviews: Water* DOI: 10.1002/wat2.1246

Li Y, Fowler HJ, Argüeso D, Blenkinsop S, Evans JP, Lenderink G, Yan X, Guerreiro SB, Lewis E, Li XF. 2020. Strong Intensification of Hourly Rainfall Extremes by Urbanization. *Geophysical Research Letters* DOI: 10.1029/2020GL088758

Liguori S, Rico-Ramirez M, Ian Cluckie. 2009. Uncertainty propagation in hydrological forecasting using ensemble rainfall forecasts. In *Proc. of Symposium JS.4 at the Joint IAHS & IAH Convention, Hyderabad*IAHS: India; 30.

Lilliefors HW. 1969. On the Kolmogorov-Smirnov Test for the Exponential Distribution with Mean Unknown. *Journal of the American Statistical Association* DOI: 10.1080/01621459.1969.10500983

Liu C, Guo L, Ye L, Zhang S, Zhao Y, Song T. 2018. A review of advances in China's flash flood early-warning system. *Natural Hazards* **92** (2): 619–634 DOI: 10.1007/s11069-018-3173-7

Liu J. 2012. Progresses and Challenges on QPE/QPF Utilization in Hydrology. *Tropical Cyclone Research and Review* **1** (2): 194–197 DOI: 10.6057/2012TCRR02.03

Liu Jun, Gao Cheng, Cui Han, Zhu Jie, Lv Gang. 2010. Research on relationship between drainage modulus and drainage area in large urbanized polder. In *2010 The 2nd International Conference on Industrial Mechatronics and Automation*IEEE; 630–633. DOI: 10.1109/ICINDMA.2010.5538227

Liu X, Lim S. 2018. An agent-based evacuation model for the 2011 Brisbane City-scale riverine flood. *Natural Hazards* DOI: 10.1007/s11069-018-3373-1

Liu Z, Todini E. 2002. Towards a comprehensive physically-based rainfall-runoff model. *Hydrol. Earth Syst. Sci.* **6** (5): 859–881

Lui D, Zhang Q. 2004. Hydraulic experiment of arc-shaped gate on the Qinhuai river. In *Environmental Hydraulics and Sustainable Water Management*, Lee J, , Lam K (eds).BALKEMA PUBLISHERS: Hong Kong; 2319.

Martina ML V, Todini E, Libralon A. 2006. A Bayesian decision approach to rainfall thresholds based flood warning. *Hydrology and Earth System Sciences Discussions* **2** (6): 2663–2706

Marty R, Zin I, Obled C. 2013. Sensitivity of hydrological ensemble forecasts to different sources and temporal resolutions of probabilistic quantitative precipitation forecasts: flash flood case studies in the Cévennes-Vivarais region (Southern France). *Hydrological Processes* **27** (1): 33–44 DOI: 10.1002/hyp.9543

Matte S, Boucher MA, Boucher V, Fortier Filion TC. 2017. Moving beyond the cost-loss ratio: Economic assessment of streamflow forecasts for a risk-Averse decision maker. *Hydrology and Earth System Sciences* **21** (6): 2967–2986 DOI: 10.5194/hess-21-2967-2017

McLuckie B. 1970. The warning system in disaster situations: A selective analysis

Mellor D, Sheffield J, O'Connell PE, Metcalfe A V. 2000. A stochastic space-time rainfall forecasting system for real time flow forecasting II: Application of SHETRAN and ARNO rainfall runoff models to the Brue catchment. *Hydrology and Earth System Sciences* DOI: 10.5194/hess-4-617-2000

Menne MJ, Durre I, Vose RS, Gleason BE, Houston TG. 2012. An overview of the global historical climatology network-daily database. *Journal of Atmospheric and Oceanic Technology* DOI: 10.1175/JTECH-D-11-00103.1

Mileti DS, Beck EM. 1975. Communication in crisis: Explaining Evacuation Symbolically. *Communication Research* **2** (1): 24–49 DOI: 10.1177/009365027500200102

Milionis N, Owen PW. 2018. Floods Directive: progress in improve and implementation need to assessing risks, while planning

Ming X, Liang Q, Xia X, Li D, Fowler HJ. 2020. Real-time flood forecasting based on a high-performance 2D hydrodynamic model and numerical weather

predictions. *Water Resources Research* DOI: 10.1029/2019wr025583

Molinari D, Menoni S, Ballio F. 2013. *Flood Early Warning Systems: Knowledge and Tools for Their Critical Assessment*. WIT Press: Southampton.

Montesarchio V, Lombardo F, Napolitano F. 2009. Rainfall thresholds and flood warning: An operative case study. *Natural Hazards and Earth System Science* DOI: 10.5194/nhess-9-135-2009

Mosavi A, Ozturk P, Chau KW. 2018. Flood prediction using machine learning models: Literature review. *Water (Switzerland)* **10** (11): 1536 DOI: 10.3390/w10111536

Nagao M, Kadoya M. 1971. Two-variate Exponential Distribution and Its Numerical Table for Engineering Application. *Bulletin of the Disaster Prevention Research Institute* **20** (3): 34

NMG. 2016. Nanjing Qinhuai River Environmental Improvement Project. Project Number: 37603. Nanjing. Available at: https://www.adb.org/projects/37603-013/main

Novack-Gottshall P, Wang S. 2019. Package 'KScorrect' Available at: https://cran.r-project.org/web/packages/KScorrect/KScorrect.pdf

O'Connell P. 2005. Interactive comment on (A Bayesian decision approach to rainfall thresholds based flood warning) by M. L. V. Martina et al. *Hydrol. Earth Syst. Sci. Discuss.* **2**

O'Connell P. 2018. Lecture on 'Technical approach to real-time flood forecasting'.School of Engineering, Newcastle University.

O'shea T, Bates P, Neal J. 2020. Testing the impact of direct and indirect flood warnings on population behaviour using an agent-based model. *Natural Hazards and Earth System Sciences* DOI: 10.5194/nhess-20-2281-2020

Olsen AS, Zhou Q, Linde JJ, Arnbjerg-Nielsen K. 2015. Comparing methods of calculating expected annual damage in urban pluvial flood risk assessments. *Water (Switzerland)* DOI: 10.3390/w7010255

Orr P, Twigger-Ross C. 2009. Communicating risk and uncertainty in flood warnings: a review of Defra/Environment Agency FCERM literature. Bristol, UK. Available at: https://www.gov.uk/government/publications/communicating-risk-and-uncertainty-in-flood-warnings-a-review-of-literature

Pagano TC, Wood AW, Ramos M-H, Cloke HL, Pappenberger F, Clark MP, Cranston M, Kavetski D, Mathevet T, Sorooshian S, et al. 2014. Challenges of Operational River Forecasting. *Journal of Hydrometeorology* DOI: 10.1175/jhm-d-13-0188.1

Pappenberger F, Cloke HL, Parker DJ, Wetterhall F, Richardson DS, Thielen J. 2015. The monetary benefit of early flood warnings in Europe. *Environmental Science & Policy* **51**: 278–291 DOI: 10.1016/J.ENVSCI.2015.04.016

Parker D. 1991. The Damage Reducing Effects of Flood Warnings. London, UK.

Parker D, Fordham M. 1996. An evaluation of flood forecasting, warning and response systems in the European Union. *Water Resources Management* **10** (4): 279–302 DOI: 10.1007/BF00508897

Parker D, Tunstall S, McCarthy S. 2007. New insights into the benefits of flood warnings: Results from a household survey in England and Wales. *Environmental Hazards* **7** (3): 193–210 DOI: 10.1016/J.ENVHAZ.2007.08.005

Parker DJ. 2017. Flood Warning Systems and Their Performance. *Oxford Research Encyclopedia of Natural Hazard Science* DOI: 10.1093/acrefore/9780199389407.013.84

Parkin G, Birkinshaw SJ, Younger PL, Rao Z, Kirk S. 2007. A numerical modelling and neural network approach to estimate the impact of groundwater abstractions on river flows. *Journal of Hydrology* DOI: 10.1016/j.jhydrol.2007.01.041

Penning-Rowsell E, Priest S, Parker D, Morris J, Tunstall S, Viavattene C, Chatterton J, Owen D. 2020. Flood and Coastal Erosion Risk Management Handbook and Data for Economic Appraisal Available at: https://www.mcm-

online.co.uk/handbook/

Petit F, Pauquet A. 1997. Bankfull Discharge Recurrence Interval in Gravel-bed Rivers. *Earth Surface Processes and Landforms* **22** (7): 685–693 Available at: https://onlinelibrary.wiley.com/doi/abs/10.1002/%28SICI%291096-9837%28199707%2922%3A7%3C685%3A%3AAID-ESP744%3E3.0.CO%3B2-J [Accessed 5 May 2021]

Pianosi F, Beven K, Freer J, Hall JW, Rougier J, Stephenson DB, Wagener T. 2016. Sensitivity analysis of environmental models: A systematic review with practical workflow. *Environmental Modelling & Software* **79**: 214–232 DOI: 10.1016/J.ENVSOFT.2016.02.008

Priest SJ, Parker DJ, Tapsell SM. 2011. Modelling the potential damage-reducing benefits of flood warnings using European cases. *Environmental Hazards* **10** (2): 101–120 DOI: 10.1080/17477891.2011.579335

Rabuffetti D, Barbero S. 2005. Operational hydro-meteorological warning and real-time flood forecasting: the Piemonte Region case study. *Hydrology and Earth System Sciences* DOI: 10.5194/hess-9-457-2005

Rai RK, van den Homberg MJC, Ghimire GP, McQuistan C. 2020. Cost-benefit analysis of flood early warning system in the Karnali River Basin of Nepal. *International Journal of Disaster Risk Reduction* DOI: 10.1016/j.ijdrr.2020.101534

Ramos M-H, Mathevet T, Thielen J, Pappenberger F. 2010. Communicating uncertainty in hydro-meteorological forecasts: mission impossible? *Meteorological Applications* **17** (2): 223–235 DOI: 10.1002/met.202

Reichel F, Verworn H-R, Krämer S, Cluckie I, Rico-Ramirez MA. 2009. Radar-based flood forecasting for river catchments. *Proceedings of the Institution of Civil Engineers - Water Management* **162** (2): 159–168 DOI: 10.1680/wama.2009.162.2.159

Ritter J, Berenguer M, Corral C, Park S, Sempere-Torres D. 2020. ReAFFIRM: Real-time Assessment of Flash Flood Impacts – a Regional high-resolution Method. *Environment International* DOI: 10.1016/j.envint.2019.105375

Robson AJ, Moore RJ, Wells S, Rudd A, Cole S., Mattingley P. 2017. Understanding the performance of flood forecasting models. Bristol, UK.

Sättele M, Bründl M, Straub D. 2015. Reliability and effectiveness of early warning systems for natural hazards: Concept and application to debris flow warning. *Reliability Engineering and System Safety* DOI: 10.1016/j.ress.2015.05.003

Sayers P, Horritt M, Penning-Rowsell E, McKenzie A. 2015. Climate Change Risk Assessment 2017: Projections of future flood risk in the UK. London, UK.

Sayers P, Lamb R, Brisley R, Hunter N, Wingfield S, Warren S, Mattingley P. 2018. Flood Standards of Protection and Risk Management Activities. Skipton.

Schröter K, Ostrowski M, Quintero F, Corral C, Velasco C, Sempere D, Hans Peter Nachtnebel B, Beyene M, Rubin C, Gocht M. 2008. Effectiveness and Efficiency of Early Warning Systems for Flash-Floods (EWASE). London, UK. Available at: http://www.crue-eranet.net/calls/ewase_finalreport.pdf

Sene K. 2008. *Flood warning, forecasting and emergency response*. Springer Science & Business Media: United Kingdom.

Sheather SJ, Jones MC. 1991. A Reliable Data-Based Bandwidth Selection Method for Kernel Density Estimation. *Journal of the Royal Statistical Society: Series B (Methodological)* DOI: 10.1111/j.2517-6161.1991.tb01857.x

Song M. 2019. SWMM model in Nanjing. In *II-Workshop of the Project 'A Virtual Collaboratory for Flood Forecasting, Flood Warning and Decision-Making Under Uncertainty in Urban Flood Management'* Newcastle.

Stanke C, Murray V, Amlôt R, Nurse J, Williams R. 2012. The effects of flooding on mental health: Outcomes and recommendations from a review of the literature. *PLoS Currents* DOI: 10.1371/4f9f1fa9c3cae

Stedinger JR. 1980. Fitting log normal distributions to hydrologic data. *Water Resources Research* DOI: 10.1029/WR016i003p00481

Stedinger JR, Vogel R, Foufoula-Georgiou E. 1993. Frequency Analysis of Extreme Events. In *Handbook of Hydrology*, Maidment D (ed.).McGraw-Hill: New York; 1424.

Van Steenbergen N, Willems P. 2014. Rainfall Uncertainty in Flood Forecasting: Belgian Case Study of Rivierbeek. *Journal of Hydrologic Engineering* DOI: 10.1061/(asce)he.1943-5584.0001004

Tanaka JS. 1987. 'How Big Is Big Enough?': Sample Size and Goodness of Fit in Structural Equation Models with Latent Variables. *Child Development* DOI: 10.1111/j.1467-8624.1987.tb03495.x

Tareghian R, Kashefipour SM. 2007. Application of fuzzy systems and artificial neural networks for flood forecasting. *Journal of Applied Sciences* **7** (22): 3451–3459 DOI: 10.3923/jas.2007.3451.3459

Thomopoulos N. 2017. *Statistical Distributions. Applications and Parameter Estimates*. Springer Nature: Switzerland. DOI: 10.1007/978-3-319-65112-5

Todini E. 1996. The ARNO rainfall-runoff model. *Journal of Hydrology* **175** (1–4): 339–382 DOI: 10.1016/S0022-1694(96)80016-3

Todini E. 2004. Role and treatment of uncertainty in real-time flood forecasting. *Hydrological Processes* **18** (14): 2743–2746 DOI: 10.1002/hyp.5687

Todini E. 2008. A model conditional processor to assess predictive uncertainty in flood forecasting. *International Journal of River Basin Management* **6** (2): 123–137 DOI: 10.1080/15715124.2008.9635342

Todini E. 2010. Predictive Uncertainty and its Use to Improve Decision-making. In *Lecture on Predictive Uncertainty, CPD Course on 'Real-Time Flood Forecasting and Warning Systems'*Newcastle University: Newcastle,UK.

Todini E. 2016. Predictive Uncertainty Assessment and Decision Making. In *Handbook of Applied Hydrology*, Singh VP (ed.).McGraw-Hill Education: USA; 26/1-26/13.

Todini E. 2017. Flood Forecasting and Decision Making in the new Millennium. Where are We? *Water Resources Management* **31** (10): 3111–3129 DOI: 10.1007/s11269-017-1693-7

USWRC. 1982. Guidelines for determining flood flow frequency

Verkade J. 2019. A global perspective on developments in operational flood forecasting systems. In *BHS Conference on Flood Forecasting* London, UK.

Verkade JS, Werner MGF. 2011. Estimating the benefits of single value and probability forecasting for flood warning. *Hydrology and Earth System Sciences* **15** (12): 3751–3765 DOI: 10.5194/hess-15-3751-2011

Wang QJ, Robertson DE, Chiew FHS. 2009. A Bayesian joint probability modeling approach for seasonal forecasting of streamflows at multiple sites. *Water Resources Research* DOI: 10.1029/2008WR007355

Wasson RJ. 2016. Uncertainty, ambiguity and adaptive flood forecasting. *Policy and Society* **35** (2): 125–136 DOI: 10.1016/j.polsoc.2016.06.002

Weerts AH, Winsemius HC, Verkade JS. 2011. Estimation of predictive hydrological uncertainty using quantile regression: Examples from the National Flood Forecasting System (England and Wales). *Hydrology and Earth System Sciences* DOI: 10.5194/hess-15-255-2011

Wei D, Urich C, Liu S, Gu S. 2018. Application of CityDrain3 in flood simulation of sponge polders: A case study of Kunshan, China. *Water (Switzerland)* DOI: 10.3390/w10040507

Williams GP. 1978. Bank-full discharge of rivers. *Water Resources Research* **16** (6)

Williams H. 1964. Human factors in warning and response systems. In *The Threat of Impeding Disaster*MIT press: Cambridge; 79–114.

Wilson EB, Worcester J. 1945. The Normal Logarithmic Transform. *The Review of Economics and Statistics* **27** (1): 17 DOI: 10.2307/1927163

WMO. 2011. Manual on Flood Forecasting and Warning. Geneva ,Switzerland.

Wu SJ, Hsu CT, Lien HC, Chang CH. 2015. Modeling the effect of uncertainties in rainfall characteristics on flash flood warning based on rainfall thresholds. *Natural Hazards* DOI: 10.1007/s11069-014-1390-2

Xuan Y, Cluckie ID, Wang Y. 2009. Uncertainty analysis of hydrological ensemble

forecasts in a distributed model utilising short-range rainfall prediction. *Hydrology and Earth System Sciences* DOI: 10.5194/hess-13-293-2009

Yang LE, Scheffran J, Süsser D, Dawson R, Chen YD. 2018. Assessment of Flood Losses with Household Responses: Agent-Based Simulation in an Urban Catchment Area. *Environmental Modeling & Assessment* **23** (4): 369–388 DOI: 10.1007/s10666-018-9597-3

Yue S. 2000. The bivariate lognormal distribution to model a multivariate flood episode. *Hydrological Processes* DOI: 10.1002/1099-1085(20001015)14:14<2575::AID-HYP115>3.0.CO;2-L

Yue S. 2001. A bivariate gamma distribution for use in multivariate flood frequency analysis. *Hydrological Processes* DOI: 10.1002/hyp.259

Zhao T, Wang QJ, Bennett JC, Robertson DE, Shao Q, Zhao J. 2015. Quantifying predictive uncertainty of streamflow forecasts based on a Bayesian joint probability model. *Journal of Hydrology* DOI: 10.1016/j.jhydrol.2015.06.043

Zhao Z, Qin Z, Cao J, Xia L. 2017. Source and Ecological Risk Characteristics of PAHs in Sediments from Qinhuai River and Xuanwu Lake, Nanjing, China. *Journal of Chemistry* DOI: 10.1155/2017/3510796

Appendix A. Additional tables

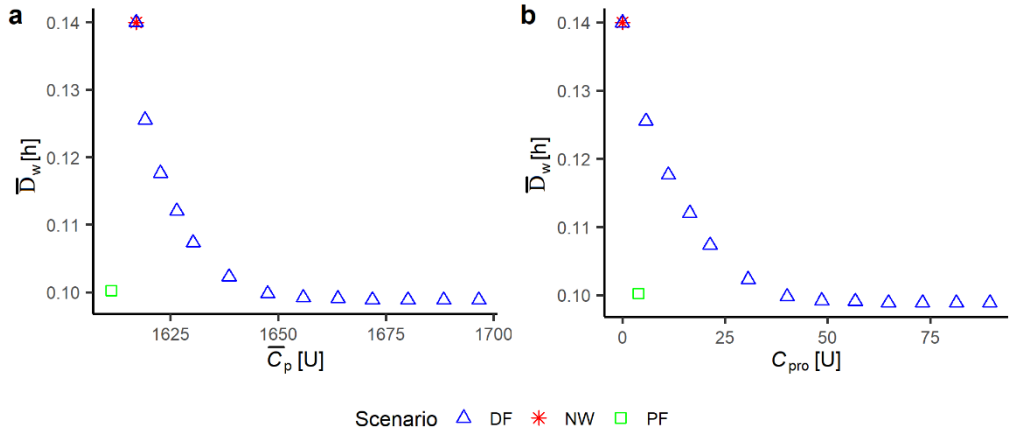
Appendix Table A-1: The pumping capacity of the pumping stations in the Shazou polder

No	Pumping station	Pumping capacity [m3/s]
1	Qingjiang Bridge	4.5
2	Dadoumen	3
3	Xiaodoumen	6
4	Liwei	15
5	Mochou	8
6	South Lake	11.55
7	City station	15
8	Xiangyang	21
9	Hujia gate	21
10	Haner	15
11	Lotus	8
12	Xingeng	24
13	Luotang	14
14	Touguan	8
15	Honggi	15.45
16	Black bridge	30
17	Zhongbao	6

Appendix Table A-2: Set of parameters defined as the best points as the Pareto front in the Probabilistic scenario (black dots in Figure 7.21)

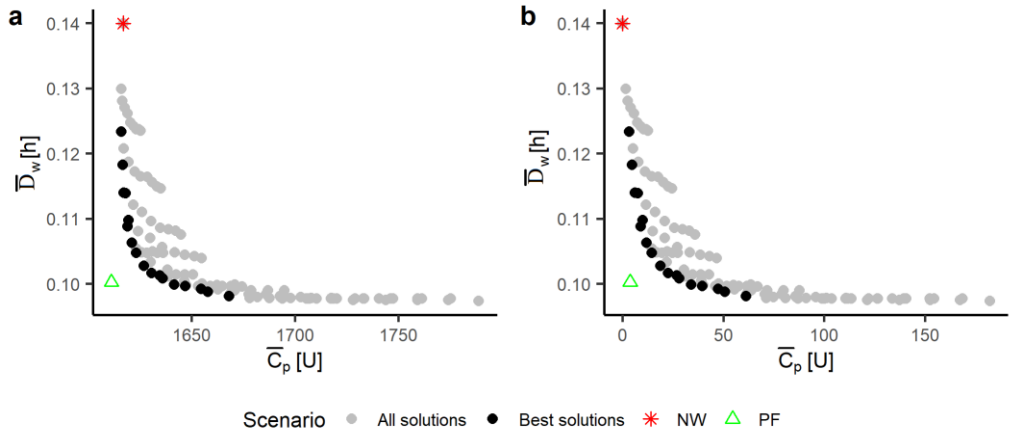
PT	***α***
0.5	0.2
0.5	0.25
0.6	0.2
0.6	0.25
0.7	0.2
0.7	0.25
0.8	0.075
0.8	0.1
0.8	0.2
0.8	0.25
0.8	0.3
0.9	0.05
0.9	0.075
0.9	0.1
0.9	0.15
0.9	0.2
0.9	0.25

Appendix B. Additional figures



Appendix Figure B-1: : Tradeoff between pumping costs and the average number of hours of waterlogging ($\bar{D}_w$) for the deterministic forecast scenario and comparison with the two benchmark scenarios

In the figure, DF: deterministic forecast; NW: no warning scenario; PF: perfect forecast. This figure shows the tradeoff between $\bar{C}_p$ and $\bar{d}_w$ and $\bar{C}_{pro}$ and $\bar{d}_w$ for the deterministic forecast scenario. The values of α considered were the following: 0, 0.025,0.05,0.075,0.10,0.15,0.2,0.25,0.30,0.35,0.40,0.45,0.5. It also shows the values of the two benchmark scenarios.



Appendix Figure B-2: Tradeoff between pumping costs and the average number of hours of waterlogging ($\bar{D}_w$) for the probabilistic forecast scenario and comparison with the two benchmark scenarios

In the figure, DF: deterministic forecast; NW: no warning scenario; PF: perfect forecast. This figure shows the tradeoff between $\bar{C}_p$ and $\bar{d}_w$ and $\bar{C}_{pro}$ and $\bar{d}_w$ for the deterministic forecast scenario. The values of α considered were the following: 0, 0.025,0.05,0.075,0.10,0.15,0.2,0.25,0.30,0.35,0.40,0.45,0.5. It also shows the values of the two benchmark scenarios.